

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5 4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: March 14, 2003, 18:44:25 ; Search time 4324 Seconds
(without alignments)
4630.604 Million cell updates/sec

Title: US-09-700-590A-22

Perfect score: 3639
Sequence: 1 MSAAESGPGTRRLNLPVMDG.....ALCHLLFAAEKETFRLRLM 688

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+.p2n.model -DEV=xlp
-Q=/cgn2.1/USPTO_spool/US9700590/runat_14032003_140630_3224/app_query.fasta_1.839
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOFCU=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=bl0sum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pct -NOM=ext -HEAPSIZE=500 -MITLEN=0 -MAXLEN=200000000
-USER=US09700590@cgn 1.1 3745 @runat 14032003 140630 3224 -NCPU=6 -ICPU=3
-NO_XLPRX -NO_MMAP -LARGEQUERIES -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.*
1: gb_ba.*
2: gb_hcg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*

29: em_vi.*
30: em_hcg_hum.*
31: em_hcg_inv.*
32: em_hcg_other.*
33: em_hcg_mus.*
34: em_hcg_pln.*
35: em_hcg_rod.*
36: em_hcg_mam.*
37: em_hcg_vrt.*
38: em_sy.*
39: em_hnco_hum.*
40: em_hnco_mus.*
41: em_hnco_other.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match Length	ID	Description
1	3583	98.5	3149	9 HSHUNKI	Y12059 H.sapiens H
2	3583	98.5	5198	9 AF386649	AF386649 Homo sapi
3	3441.5	94.6	2727	10 AF461396	AF461396 Mus muscu
4	3436.5	94.4	5281	10 AF273217	AF273217 Mus muscu
5	3436.5	94.4	6000	10 AF461395	AF461395 Mus muscu
6	2556.5	70.3	2199	10 BC008532	BC008532 Mus muscu
7	1931	53.1	3033	10 BC031536	BC031536 Mus muscu
8	1928	53.0	2933	10 AF269193	AF269193 Mus muscu
9	1922	52.8	3028	9 HUMOREX	D26362 Human mRNA
10	1901	52.2	3649	10 AF045462	AF045462 Mus muscu
11	1888	51.9	4053	9 HSEFSHR	X62083 H.sapiens m
12	1888	51.9	4053	9 HUMFSHG	M80613 Human homol
13	1888	51.9	4664	6 AA409010	AA409010 Sequence
14	1888	51.9	4664	6 HUMKIA9001	D42040 Human mRNA
15	1877.5	51.6	2548	9 BC032124	BC032124 Homo sapi
16	1837	50.5	4667	9 HSM804033	AF132722 Homo sapi
17	1684	46.3	2522	5 AF191032	AF191032 Myxine gl
18	1573	43.2	37644	6 A95302	A95302 Sequence 33
19	1573	43.2	92863	12 GGBLOCUS	AL023516 Gallus ga
20	1548.5	42.6	4651	5 GGRING3GE	X96669 G.gallus RI
21	1540.5	42.3	3429	9 BC011898	BC011898 Homo sapi
22	1540	42.3	3391	10 AF358660	AF358660 Mus muscu
23	1535.5	42.2	2350	9 BC000477	BC000477 Homo sapi
24	1520	41.8	1657	10 AB010247	AB010247 Mus muscu
25	1520	41.8	1688	10 AB010246	AB010246 Mus muscu
26	1501.5	41.3	3106	9 AF019085	AF019085 Homo sapi
27	1500	41.2	1554	10 AB010248	AB010248 Mus muscu
28	1458.5	40.1	3404	9 AK056504	AK056504 Homo sapi
29	1451.5	39.9	11493	10 MUI1KLT	AL009226 Mus muscu
30	1451.5	39.9	273800	10 AF100956	AF100956 Mus muscu
31	1450	39.8	1884	9 AK024312	AK024312 Homo sapi
32	1441	39.6	10222	10 D89801	D89801 Mus muscu
33	1441	39.6	216734	5 AB073376	AB073376 Oryzias l
34	1440.5	39.6	274856	2 AC098547	AC098547 Rattus no
35	1401	38.5	2206	10 AF318183	AF318183 Mus muscu
36	1380.5	37.9	5632	3 DROFSHB	M23222 D.melanogas
37	1380.5	37.9	7792	3 DROFSHA	M23221 D.melanogas
38	1237.5	34.0	1741	9 S78771	S78771 NAT=CPG 181
39	1234	33.9	37518	2 AC014003	AC014003 Drosophi
40	1234	33.9	181178	3 AC023713	AC023713 Drosophi
41	1234	33.9	306267	3 AE003442	AE003442 Drosophi
42	1211.5	33.3	14561	9 HSRING3GE	X96670 H.sapiens R
43	1211.5	33.3	38902	9 HSO27	Z96104 Human DNA s
44	1211.5	33.3	40740	9 HSO14	Z84497 Human DNA s
45	1200.5	33.0	98309	2 AL671924	AL671924 Homo sapi

RESULT 1

ALIGNMENTS

[illegible]

Db	403	AACCACTGCAGAACTCGGCTCAGAGTGCGTCTCAAGACACTATGGAAACACCAAGTTTGA	462
Qy	81	TrpProheGInGInProValAspAlaValIysIleAsnLeuProAspTyrTyrIysIle	100
Db	463	TGGCGCTTTCAGAGCGCTGGAGTGGCGGCAAGCTGAACCTCCGATCTATTAAGATC	522
Qy	101	IleIysThrProMetAspMetGlyThrIleIysIysArgLeuGluAsnAspTyrTyr	120
Db	523	ATTMAACGCGCTTATGATATGGACAAATMAAGAAAGCGCTTGGAAACACTATTATTCGG	582
Qy	121	AsnAlaGInGluCysIleGlnAspPheAsnThrMetPheThrAsnCysTyrIleTyrAsn	140
Db	583	AATGCTCAGGAATGTATTCAGAGACTTCAACACTATGTTTACAAATGTTTACATCTACAC	642
Qy	141	LysProGlyAspAspIleValLeuMetAlaGluAlaLeuGluIysLeuPheLeuGlnIlys	160
Db	643	AAGCTCGAGATGACATAGCTCTTAAGGAGAAAGCTCTGGAAAGACTCTTTGCCAAAA	702
Qy	161	IleAsnGluLeuProThrGluGluThrGluIleMetIleValGlnAlaIysGlyArgGly	180
Db	703	ATAAATAGTCACTCCACAGAAAGAACCGAGATATGATATGTCAGAGCAAAAGAGAGGA	762
Qy	181	ArgGlyArgIysGluThrGlyThrAlaIysProGlyValSerThrValProAsnThrThr	200
Db	763	CGTGGAGGAAAGAAACAGGACAGCAAAACCTGGGGTTTCCAGGGTTCACAAACAAC	822
Qy	201	GlnAlaSerThrProProGInThrGInThrProGInProAsnProProProValGlnAla	220
Db	823	CAACATCTGACTCTCCGCGAGACCAGACCCCTCAGCCGAATCTCTCCGTGTGAGGCC	882
Qy	221	ThrProHisProPheProAlaValThrProAspLeuIleValGlnThrProValMetThr	240
Db	883	ACGCTCAACCCCTTCCCTGCGTACCCTCGAGACTATGTCTCAAGCCCTGTCTATGACA	942
Qy	241	ValValProProGInProLeuGInThrProProProValProProGInProGInProPro	260
Db	943	GTGTGGCTCCCCAGACCACTGAGAGAGCCCGCCAGTGGCCCCCCAGCAACCCCA	1002
Qy	261	ProAlaProAlaProGInProValGInSerHisProProIleIleAlaAlaThrProGIn	280
Db	1003	CCCGCTCAGGCTCCCGACCCGTACAGACCAACCATCATGTGGCGCACCCACAG	1062
Qy	281	ProValIysThrIysIysGlyValIysArgIysAlaAspThrThrThrProThrThrIle	300
Db	1063	CCTGTGAAGAAAGAGGAGAGAGAGAAAGCAAGCAACCAACCCACCAACCACTT	1122
Qy	301	AspProIleHisGluProProSerLeuProProGluProIlyThrThrIysLeuGlyGln	320
Db	1123	GACCCCATTCAGAGCAACCTCTGCTGCCCGAGCCCAACCAACCAAGTGGCGAG	1182
Qy	321	ArgArgGluSerSerArgProValIysProProIlyIysAspValProAspSerGlnGln	340
Db	1183	CGGCGGAGAGACACGCGCTGTGAACCTCCAAGAGAGAGTCCCGACTCTCAGAG	1242
Qy	341	HisProAlaProGluIlyIysSerSerIysValSerGluGlnLeuIysCysCysSerGlyIle	360
Db	1243	CACCCAGACCAAGAGACCAACAGGCTCTGGAGCACTAAATGCTGCAGCGGCATC	1302
Qy	361	LeuIysGluMetPheAlaIysIysHisAlaIleTyrAlaThrProPheTyrIysProVal	380
Db	1303	CTCAGAGAGATGTTTGCAGAAAGACACGCGGCTCAGCTGGCCCTTCTACAGCGCTGTG	1362
Qy	381	AspValGluAlaLeuGlyLeuHisAspTyrCysAspIleIleIysHisProMetAspMet	400
Db	1363	GACGTGAGAGCACTGGGCTTACACAGACTACTGTGATCATATCAACACCCCATGGACATG	1422
Qy	401	SerThrIleIysSerIysLeuGluAlaArgGluTyrArgAspAlaGlnGluPheGlyAla	420
Db	1423	AGCAACAATCAAGTCTAACTCGAGAGCCCGGTAGTACCGTGAAGTCCAGAGATTGTGCT	1482
Qy	421	AspValAlaGluMetPheSerAsnCysTyrIysTyrAsnProProAspHisGluValVal	440
Db	1483	GACGTCGATTTGATGTTCTCACTGCTATAGTAAACCCCTCCGACCAATGAGGTGGTG	1542

QY 81 TrpProheGInGInProValAspAlaValLysLeuAenLeuProAspTyrTyrLysIle 100
 Db 463 TGGCTTTCCAGACACCTGTCGATGCCCTCAAGTCGAACCTCCGATCTACTATTAAGTC 522
 QY 101 TleYsthrProMetAspMetGlyThrTleYsthrLysArgLeuGluAsnAsnTyrTyrTrp 120
 Db 523 ATTAATAAGCTTATGATATGGAACAATAAAGAGCCCTTGGAAGAACACTATTAATACGG 582
 QY 121 AsnAlaGInGInCysTleGlnAspPheAsnThrMetPheThrAsnCysTyrTleTyrAsn 140
 Db 583 AATGCTCAGAGATGATATCAGAGACTTCAACACTATGTTTACAAATGTTTACATCTACAC 642
 QY 141 LysProGInLysAspAspIleValLeuMetAlaGluAlaLeuGluLysLeuPheLeuGlnLys 160
 Db 643 AAGCTGAGATGATCATGTGCTTAATGCAAGAGCTCTGGAAGAACCTTCTTGCAAAA 702
 QY 161 TleAsnGluLeuProThrGluGluThrGluTleMetTleValGlnAlaLysGlyValArgGly 180
 Db 703 ATTAATAGTATACCAACAGAGAAACCGAGATCATGATAGTCCAGCAAAAGAGAGAG 762
 QY 181 ArgGlyArgLysGluThrGlyThrAlaLysProGInLysValSerThrValProAsnThrThr 200
 Db 763 CGTGGAGAGAAAGAAACAGAGACAGCAAAACCTGGCGTTCCAGCGTACCAAAACCAACT 822
 QY 201 GlnAlaSerThrProProGInThrGlnThrProGInProAsnProProProValGlnAla 220
 Db 823 CAAGCATGATCTCTCCGACAGACCCAGACCCCTCAGCCGATCTCTCTGTGAGAGCC 882
 QY 221 ThrProHisProPheProAlaValThrProAspLeuIleValGlnThrProValMetThr 240
 Db 883 ACGCTCAGCCCTTCCGCGTCACCCCGGACCTCATGTGTCAGAACCCCTGTCTCATGACA 942
 QY 241 ValValProProGInProLeuGlnThrProProProValProProGInProGInProPro 260
 Db 943 GTGGGCTCTCCAGACCTGACAGCGCCCGCGCAGTGGCCCGCCAGCCCAACCCCA 1002
 QY 261 ProAlaProAlaProGInProValGlnSerHisProProIleIleAlaIleThrProGln 280
 Db 1003 CCCGCTCAGCTCCCAACCCGCTACAGAGCCACCAACCATCATGTGGCCACCCACAG 1062
 QY 281 ProValLysThrLysLysGlyValLysArgLysAlaAspThrThrThrProThrThrIle 300
 Db 1063 CCTGGAAGACAAAGAAAGAGTGAAGAGAGAAAGACACACCAACCCCAACCCCAATT 1122
 QY 301 AspProIleHisGluProProSerLeuProProGluProLysThrThrLysLeuGlnLys 320
 Db 1123 GACCCCATTCAGAGCCACCTCGCTGCCCGGAGCCCAAGACCAACCAAGCTGGGCGAG 1182
 QY 321 ArgArgGluSerSerArgProValLysProProLysLysAspValProAspSerGlnGln 340
 Db 1183 CGCGCGAGAGACAGCCGCGCTGTGAACCTCCAAAGAAAGAGCTCCGACTCTACAGAG 1242
 QY 341 HisProAlaProGluLysSerSerLysValSerGlnGlnLeuLysCysCysSerGlyIle 360
 Db 1243 CACCAGACAGCAGAGAGAGGAGGAGGCTCTGAGAGAGCTCAAGTGTGAGCGGAGATC 1302
 QY 361 LeuLysGluMetPheAlaLysLysHisAlaAlaIleTyrAlaIleTyrProPheTyrLysProVal 380
 Db 1303 CTCAAGGAGATGTTTCCAAAGAGACAGCGCGCTCAGCTGCGCTTCTCAAGAGCTGTG 1362
 QY 381 AspValGluAlaLeuGlyLysHisAspTyrCysAspIleIleLysHisProMetAspMet 400
 Db 1363 GACGTGGAGGAGCTGGGCTTACACAGACTCTGTGACATCATTAAGACACCCCAATGAGCATG 1422
 QY 401 SerThrIleLysSerLysLeuGlnAlaArgGluTyrArgAspAlaGlnGluPheGlyAla 420
 Db 1423 AGCAACAATCAAGTCTAAATCGAGGCCCTGTGATACCGTGAGCTCAGAGAGTTTGTTGTT 1482
 QY 421 AspValAlaGluMetPheSerAsnCysTyrLysTyrAsnProProAspHisGluValVal 440
 Db 1483 GACGCTCAGTATGATGTTCTCCAACTGATATAAGTACAAACCTCTGAGCATGAGGTGTG 1542

QY 441 AlaMetAlaArgLysLeuGlnAspValPheGluMetArgPheAlaLysMetProAspGlu 460
 Db 1543 GCCATGGCCCGCAAGCTCCAGAGATGTTGTAATATGCGCTTGTCCAGAGTCCCGAGAG 1602
 QY 461 ProGluGluProValValAlaValSerSerProAlaValProProProThrLysValVal 480
 Db 1603 CTTAGAGAGCAGTGTGGCGGTCTCTCCCGCAGTGTGCCCTCCACCAAGATTGTG 1662
 QY 481 AlaProProSerSerSerAspSerSerSerSerSerSerSerSerSerSerSerSerThr 500
 Db 1663 GCCCGCCCTCATCAGACAGACAGCAGCAGATAGCTCCTCGGACATGACAGTTGACT 1722
 QY 501 AspAspSerGluGluGluAlaGlnArgLeuAlaGluLeuGlnGlnGlnLysValAla 520
 Db 1723 GATGACTCTGAGAGAGAGCAGCCAGCGGCTGAGCTGAGCTCCAGAGAGAGCTCAAGGCC 1782
 QY 521 ValHisGluGlnLeuAlaAlaLeuSerGlnProGlnGlnAsnLysProLysLysGlu 540
 Db 1783 GTGCACAGCAGCTTGGAGCCCTCTTCACCCCGACAGAAACCAACCAAGAAAGAGAG 1842
 QY 541 LysAspLysLysGluLysLysLysGluLysLysLysArgLysGluGluValGluGluAsn 560
 Db 1843 AAAGACAAAGAT 1902
 QY 561 LysLysSerLysAlaLysGluProProProLysLysThrLysLysAsnAsnSerSerAsn 580
 Db 1903 AAAAAAGCAAGCCAAAGAACTCTCTTAAAAAGCAAGAAATAATAGCAGCAAC 1962
 QY 581 SerAsnValSerLysLysGluProAlaProMetLysSerLysProProProThrTyrGlu 600
 Db 1963 AGCAATGTGAGCAAG 2022
 QY 601 SerGluGluGluAspLysCysLysProMetSerTyrGluGluLysArgGlnLeuSerLeu 620
 Db 2023 TCGAGAGAGAGAGCAAGTGCAGAGCTTATGTCTTATAGAGAGAGAGAGAGAGAG 2082
 QY 621 AspIleAsnLysLeuProGluGluLysLeuGluValArgValHisIleIleGlnSerArg 640
 Db 2083 GACATCAACAAAGCTCCCGGAGAGAACTGGCGCGCTGTGGTGCACATCATCACTACGCG 2142
 QY 641 GluProSerLeuLysAsnSerAsnProAspGluIleGluIleAspPheGluThrLeuLys 660
 Db 2143 GAGCCCTCCCTGAAAGATTCACACCCGACAGATTAATAATCGACTTTGAGACCTTGAAG 2202
 QY 661 ProSerThrLeuArgGluLeuGluValAlaLeuCysHisLeuLeuPheAlaGluGluLysGlu 680
 Db 2203 CCGTCCACACTGCGTGAAGCT-GGAGCCCTATGTCACTCTCTGTGGCGAAGAAAGAA 2261
 QY 681 ThrPheLysLeuArgLysLeuMet 688
 Db 2262 ACCTC-AAAGTGAGAAAGTTGATG 2284
 RESULT 3
 AF461396 2727 bp mRNA linear ROD 13-MAY-2002
 LOCUS
 DEFINITION
 Mmusculus bromodomain-containing protein BRD4 short variant
 (Brd4) mRNA, complete cds.
 AF461396
 VERSION
 AF461396.1 GI:18308126
 SOURCE
 ORGANISM
 Mus musculus.
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE
 1 (bases 1 to 2727)
 Houzelstein, D., Bullock, S.L., Lynch, D.E., Grigorieva, E.F.,
 Wilson, V.A. and Bedington, R.S.
 Growth and early postimplantation defects in mice deficient for the
 bromodomain-containing protein Brd4
 Mol. Cell. Biol. 22 (11), 3794-3802 (2002)
 JOURNAL
 MEDLINE
 PUBMED
 21993140
 11997514
 REFERENCE
 2 (bases 1 to 2727)

AUTHORS Houzelstein, D.
TITLE Direct Submission
JOURNAL Submitted (19-DEC-2001) Genetique et Developpement, Institut Jacques Monod, 2 Place Jussieu, Paris 75 251, France

FEATURES
 source Location/Qualifiers

1..2727
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /chromosome="17"
 /map="between D17Mit181 and D17Mit191"
 1..2727
 /gene="Brd4"
 242..2413
 /gene="Brd4"
 /codon_start=1
 /product="bromodomain-containing protein BRD4 short variant"

gene
 CDS

/protein_id="AA167834.1"
 /db_xref="GI:18308127"

/translation="MSTESGPGTRRLNLFVMDGLTSGMSTTQAQAPOPANMASTN
 PPPPTSNPNKPKRQTNQOYLRLVVLKTLMKHOFAMPPOQPDVAIKLIPYKLIK
 TMDMGTLKRLRNNYNNAEQIODFNTMPTNICYVNRBGDDIVLMARLKLFLQK
 INELPTEETEMITVQAKGRGRKRTGTAKPGVSPVPTQASTSPQOTPOONRPP
 VQATTHPPAVTIDLTADPPVMTWPPQQLQTPSPVPPOPPPPAPFPQVSHPEPI
 ATTPQPVTKKVKRKAADTTPTTIDPIHEPSLAEPYAKLGPRESSREVKKPK
 DVPDSQOHGPKSKISSEOLKCCSGILKEMFAKKAAYAMPYKVDVALGLHDYC
 DIKHPMDSTIKSLKSEKREVDAGFADVRLPSSSDSSSSSSSDSTDSSEER
 FEMRFPKMPDEPEPVTVSPVAPVPTKVA PPSSDSSSSSSSDSTDSSEER
 AORLAELQQLKAVNHOQLAALSOPOONRKKKKKKKKKKKKKKKKKKKKKKKK
 KALPKRTKKNNSNVSKKEPVPTKTPPTPESEEDKKPKMSYEKKRLSDIN
 KLPGELGKRVVHI IQSREPSLKNSPDELDIEFLKLPSTLELELYVTSCLKRRKP
 QAEKVDVIVGSSKMGKFSSESSESTSESSSDSESETGPA"

BASE COUNT 802 a 763 c 636 g 526 t
ORIGIN

Alignment Scores:

Pred. No.: 1,02e-132 Length: 2727
 Score: 3441.50 Matches: 659
 Percent Similarity: 96.80% Conservative: 7
 Best Local Similarity: 95.78% Mismatches: 21
 Query Match: 94.57% Indels: 3
 DB: 10 Gaps: 1

US-09-700-590A-22 (1-688) x AF461396 (1-2727)

QY 1 MetSerAlaGluSerGlyProGlyThrArgLeuArgAsnLeuProValMetGlyAspGly 20
 DB 242 ATGCTTACGAGAGCGCCCTGGGACAAAGATTGACAGTAATGCGAGTGAATGAGGAGATGGA 301
 QY 21 LeuGluThrSerGlnMetSerThrThrGlnAlaGlnAlaGlnProGlnProAlaAsnAla 40
 DB 302 CTGGAACCTCCCAATGTCTTACACGACGAGCCCAAGCCCAAGCCCAAGCAATGCA 351
 QY 41 AlaSerThrAsnProProProGluThrSerAsnProAsnLysProLysArgGlnThr 60
 DB 362 GCCAGACCAATCTCTCCACCCCAAGAGACCTCCAACTTAACAAGCCCAAGAGACAGACA 421
 QY 61 AsnGluLeuGlnThrLeuLeuArgValValLeuLysThrLeuTrpLysHisGlnPheAla 80
 DB 422 AACCAACTGCAATATCTGCTCAGAGTGGTCTCAAGACACTATGAGAAACCAAGTTTGGC 481
 QY 81 TrpProPheGlnGlnProValAspAlaValLysLeuAsnLeuProAspTyrTyrLysIle 100
 DB 482 TGGCCTTTCCAGAGAGCCCGGATGCGGTCAAGCTGAACCTCCCTGATTTACTTAAGATT 541
 QY 101 IleLysThrProMetAspMetGlyThrIleLysLysArgLeuLysAsnAsnTyrTyrTrp 120
 DB 542 ATTAATAACACCCATGATATGAGAAACAATAAGAAGCGCTTGAATAACAATATTACTGG 601
 QY 121 AsnAlaGlnGluCysIleGlnAspPheAsnThrMetPheThrAsnGlyTyrIleTyrAsn 140
 DB 602 AATGCTCAGGAATGATCCAGACTTCAACACTATGTTTCAAAATGTTTACATCTATAAC 661

QY 141 LysProGlyAspAspIleValLeuMetAlaGluAlaLeuGlnLysLeuPheLeuGlnLys 160
 DB 662 AAGCCCTGAGATGACATCTCTTAATGAGAGAGCTCTGGAGAACTCTTCTTGCAGAAAA 721
 QY 161 IleAsnGluLeuProThrGlnGluThrGlnIleMetIleValGlnAlaLysGlyArgGly 180
 DB 722 ATCAATGAACCTGCTACAGAAAGAACTGAGATCATATGATTCAGGCAAGAGAGAGGA 781
 QY 181 ArgGlyArgLysGlnThrGlnIleThrAlaLysProGlyValSerThrValProAsnThrThr 200
 DB 782 CAGAGAGAGAAAGAAACAGAGACAGCAAGACCTGCTGTATCCACGGTACCAGCAACAAC 841
 QY 201 GlnAlaSerThrProProGlnThrGlnThrProGlnProAsn--ProProProValGln 219
 DB 842 CAAGCATCAACTTCTCCGAGACCCAGACGCTCAGCAACAACCTCTCCACTGTGCAG 901
 QY 220 AlaThrProHisProPheProAlaValThrProAsnLeuIleValGlnThrProValMet 239
 DB 902 GCCACAACCTCACCCCTTCTCTGCTGCACCCAGACCTTATGCCCAGCCTCTGTCAATG 961
 QY 240 ThrValAlaProProGlnProLeuGlnThrProProProValProProGlnProGlnPro 259
 DB 962 ACAATGTGCCCCCTTACAGCACTTCCAGACTCTTCCACGGTACCCCAAGCCACCAACC 1021
 QY 260 ProProAlaProAlaProGlnProValGlnSerHisProProIleIleAlaAlaThrPro 279
 DB 1022 CCACCTGCTCCAGTTCCACAGCCTGTGCAGAGTCAACCCGCCATCATTCGACCAACCCC 1081
 QY 280 GlnProValLysThrLysGlyValLysArgLysAlaAspThrThrThrProThrThr 299
 DB 1082 CAGCCCTGTAAGACAAAGAAAGGGGAGAGAGAAAGAGATACCAACCCCTTACACAC 1141
 QY 300 IleAspProIleHisGluProProSerLeuLeuProProGlnProLysThrThrLysLeuGly 319
 DB 1142 ATGACACCCATTCATAGGACCACTCTGACCTGCGCCCAAGCCCAAGCCCAAGCTGGGT 1201
 QY 320 GlnArgArgGluSerSerArgProValLysProProLysLysAspValProAspSerGln 339
 DB 1202 CCTCGCGGAGAGAGAGAGACCTGTGAGACCTCCAAAGAGATGTACCGGACTCACAG 1261
 QY 340 GlnHisProAlaProGlnLysSerSerLysValSerGlnLysLeuLysCysCysSerGly 359
 DB 1262 CAGCACCCAGGGGCAGAGAGAGAGAGAGAGATCTCGACACACTTAAGTGTGCACTGTC 1321
 QY 360 IleLeuLysGluMetPheAlaLysLysHisAlaAlaTyrAlaTyrProPheTyrLysPro 379
 DB 1322 ATCTTAAGAGAGATGTTTCTCCAAAGAAACATGCTGCTATGCTTCTTCTTCAACGCT 1381
 QY 380 ValAspValGluAlaLeuGlnLysLeuHisAspTyrCysAspIleIleLysHisProMetAsp 399
 DB 1382 GTGGATGTGAGAGCACTGGGTCTGCACGACTACTGTGACATCATCAACATCCATGGAC 1441
 QY 400 MetSerThrIleLysSerLysLeuGlnAlaArgGluTyrArgAspAlaGlnGluPheGly 419
 DB 1442 ATGAGACACATCAAGCTTAACCTAGAGTCCGAGAGTCCAGAGATGCCAGAGAAATTTGAT 1501
 QY 420 AlaAspValArgLeuMetPheSerAsnGlyTyrLysTyrAsnProProAspHisGluVal 439
 DB 1502 GCTGATGTCGATGATGATGTTCTTCCACGCTGCTACAGATACAAACCCCTTACATGAAG 1561
 QY 440 ValAlaMetAlaArgLysLeuGlnAspValPheGluMetArgPheAlaLysMetProAsp 459
 DB 1562 GTAGCCATGGCTGAGAAATCCAGAGATGTTTGAATCGCTTGGCCAAAGTGGCTGAT 1621
 QY 460 GluProGluGluProValValAlaValSerSerProAlaValProProProThrLysVal 479
 DB 1622 GAGCCTGAAGAGACCAAGTTGTTACAGTGTCTCTCTGCAAGTCCACCCCTTCAAAAGTG 1681
 QY 480 ValAlaProProSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 499
 DB 1682 GTAGCCCAACCCCTCATGTATGATGACAGACAGACGACAGAGTCTTCCAGACGACAGTTC 1741
 QY 500 ThrAspAspSerGluGluGluArgAlaGlnArgLeuAlaGluLeuGlnGluGlnLeuLys 519


```

Db 601 AACCTGGAGATGACATCGCTTAATGCAAGCTCTGAGAAAGCTCTTTCGCAAAAA 660
Qy 116asnlguleupProthrGlugluThrGlul1ewet11eVal1Gln1alysGl1ArgGl1 180
Db 661 ATCAATCAATCTGCTTACAGAAAGAACTGAGATCATGTATGCTCCAGGAAAGAAAGAGA 720
Qy 181 ArgGl1YArg1ysgluThrGl1YThr1alysProGl1YVal1SerThrVal1ProasnThrThr 200
Db 721 CGAGGGAGAAAGAAACAGAGGGGCGCAAGCCTGTGTATCTCAAGGTACCAAAACAACT 780
Qy 201 Gln1alaserThrProProGlnThrGlnThrProGlnProasn--ProProProVal1Gln 219
Db 781 CAAGCATCAACTTCTCGCAGAGCCAGACGCTCAGAGAAACCTCTCCACCTGTCAG 840
Qy 220 AlAThrProHisProPheProAlaVal1ThrProAspleu11eVal1GlnThrProVal1Met 239
Db 841 GCCACACTACACCCCTTCTGCTGTCAACCCAGACCTTATGGCCAGCCTCTGTATG 900
Qy 240 ThrVal1alProProGlnProLeuGlnThrProProProVal1ProProGlnProGlnPro 259
Db 901 ACAATGTGCCCCCTCAGCAGCTTCAAGCTCTTCACTGAGTACCCGCCAGCCACCAACC 960
Qy 260 ProProAlaProAlaProGlnProVal1GlnSerHisProPro11e1al1al1athrPro 279
Db 961 CCACCTGCTCCAGTTCACAGCTGTGACAGATCACCCGCCCATCATTTGCGACCAACCCC 1020
Qy 280 GlnProVal1SerThr1ys1ysgl1YVal1LysArg1Lys1alAspThr1Thr1ThrProThr 299
Db 1021 CAGCTGTGAAGCAAAAGAAAGGAGTGAAGAGAAAGCAATACACACACCCCTACACCC 1080
Qy 300 11eAspPro11eHisgluProProSerLeuProProGluPro1YThr1Thr1LysLeuGl1 319
Db 1081 ATGACCCCCATTATAGAGCAACCTCTGCCCCAGGCCAAAGCCGCCAGAGCTGGGT 1140
Qy 320 GlnArgArgLysSerSerArgProVal1LysProPro1YLysAspVal1ProAspSerGln 339
Db 1141 CCTCGGGGAGAGACAGACAGCTGTGAAAGCTTCAAAAGAGATGTAACCGGACTCAG 1200
Qy 340 GlnHisProAlaProGlu1YysSerSer1Val1SerGln1Leu1YysCysCysSerGly 359
Db 1201 CACACCCAGGAGGCGCAGAGAAAGACAGCAAGATCTCTGAGAGCTTAAAGTGTCTCAGTGC 1260
Qy 360 11eLeuLysgluMetPheAlaLysLysHisAlaAla1Yz1al1at1rProPhe1YThrPro 379
Db 1261 ATCTCAAGAGAGATGTTTGGCAAGAAACATGCTGCTATGCTTGGCTTCTTCAAGACCT 1320
Qy 380 ValAspVal1Glu1alLeuGlyLeuHisAspTyrCysAsp11e11eLysHis1sProme1Asp 399
Db 1321 GTGATGTGAGGACACTGGGTCTGACAGCTACTGTGACATCAATCAATCCCATGAGAC 1380
Qy 400 MetSerThr11eLysSer1YsLeuGlu1alArgLyu1YzArgAsp1alGln1uPheGly 419
Db 1381 ATAGAGACATCAAGCTTAACTTAAGAGTCCCGAGAGTACAGAAATGCCCGAAATTTGGT 1440
Qy 420 AlaAspVal1ArgLeuMetPheSerAsnCysTyr1YsTyrAsnProProAspHisgluVal 439
Db 1441 GCTGATGTCGATTTGATTTCTCCAACTGCTTCAAGTCAACCCCCCTTGACATGAGTG 1500
Qy 440 Val1alMet1alArgLysLeuGlnAspVal1PheGluMetArgPheAlaLysMetProAsp 459
Db 1501 GTAGCCATGCTCGAAACTCCAGAGATGTGTGAATGCGCTTGGCAAGATGCGCTGAT 1560
Qy 460 GluProGluGluProVal1alVal1alVal1SerSerProAla1alProProProThr1YsVal 479
Db 1561 GAGCCCTGAAGAGCAGTGTTCAGTGTCTCTCTGACATGCGACCCCTTACAAAGGTG 1620
Qy 480 Val1alProProSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 499
Db 1621 GTAGCCCCACCTCATCTTATGTACAGACAGACAGACAGTTCCTTCCAGACCGACAGTTC 1680
Qy 500 ThrAspAspSer1uGluGluArg1alGlnArgLeu1alGluLeuGlnGlnGlnLeuLys 519

```

```

Db 1681 ACTGACGACTCTGAGCAAGAGCCAGCCGCTGCTGAACTCCAGAAACGCTCAAG 1740
Qy 520 AlaVal1HisgluGlnLeu1al1alLeuSerGlnProGlnGlnAsnLysProLysLys 539
Db 1741 GCCGTGCATGAGGAGCTTCCAGCCCTCTCAGAGCCCCACAGACAGAAACCAAGAAAG 1800
Qy 540 GluLysAsp1YsLysgluLysLysLysgluLysHisLysArgLysgluGluVal1GluGlu 559
Db 1801 GAGAGAGACAAAGAAAGAAAGAAAGAAAGCAAAAGCAAAAGAAAGAGTGGAGAA 1860
Qy 560 AsnLysLysSer1YsLysgluProProProProLysLysThrLysLysAsnSerSer 579
Db 1861 AATPAAAAAAGCAAAACCAAGAACTTCTCCAAAAAAGCAAAAGAAATAACAGCAC 1920
Qy 580 AsnSerAsnVal1SerLysLysgluProAlaPromeLysSerLysProProThrTyr 599
Db 1921 AACAGCAATGTAGCAAGAAAGAACAGTACCCAGCAAGACCAAGCCGCTCCACATAT 1980
Qy 600 GluSerGluGluGluAspLysCysLysPromeSerTyrGluGluLysArgGlnLeuSer 619
Db 1981 GAATCAGAAAGAGAGATAGTAAAGCCCATGTCTTATGAGGAGAAAGCGCAGTAAAT 2040
Qy 620 LeuAsp11eAsnLysLeuProGluGluLysLeuGluArgVal1His11eGlnSer 639
Db 2041 CTAGATATCAACAACTTCTGCTGAGAAAGTAAAGCCGCTGATACACATTAATTCAGTCA 2100
Qy 640 ArgGluProSerLeuLysAsnSerAsnProAspGlu11eGlu11eAspPheGluThrLeu 659
Db 2101 AGGGAACCATCACTTAAATCTCCAAACCCGATGAGATGAGATTGACTTTGAGACCTTG 2160
Qy 660 LysProSerThrLeuArgLysLeuGlu1YalLeuCysHisLeuLeuPheAlaGluGluLys 679
Db 2161 AAGCATCTTACACTACAGAGATGTT-GGAGCGATATGTCACTCTGTTTCCGGAAGAAAG 2219
Qy 680 GluThrPheLysLeuArgLysLeu 687
Db 2220 GAAACCTC-AAAGTAAAAAGTTG 2242

```

RESULT 5
AF461395
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source

AF461395 6000 bp mRNA linear ROD 13-MAY-2002
Mus musculus bromodomain-containing protein BRD4 long variant
(Brd4) mRNA, complete cds.
AF461395.1 GI:18308124

Mus. musculus.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 6000)
Houzelstein,D., Bullock,S.L., Lynch,D.E., Grigorieva,E.F.,
Wilson,V.A. and Beddington,R.S.
Growth and early postimplantation defects in mice deficient for the
bromodomain-containing protein Brd4
Mol. Cell. Biol. 22 (11), 3794-3802 (2002)

11997514
2 (bases 1 to 6000)
Houzelstein,D.
Direct Submission
Submitted (19-DEC-2001) Genetique et Developpement, Institut
Jacques Monod, 2 Place Jussieu, Paris 75 251, France
Location/Qualifiers
1. 6000
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="17"
/map="between D1Mit181 and D1Mit191"
1. 6000
/gene="Brd4"
242..4444
/gene="Brd4"

```

/codon_start=1
/product="Bromodomain-containing protein BRD4 long
variant"
/db_xref="GI:18308125"
/translation="WSTESGEGTRLRNLPMVNGCLETSQNSTTQAOAPOPANASIN
PPEPESNPNKPKQTNQQLQLRVLTLMKQFAPFOQPDVAVLNPDYKLIK
TPMDMGTIKRLENNYWNAAQECIODFTNTCYINKGQDILVMAEALKEFLQK
INELPTEETIMIVQAKGRGRKETGAAGKGVSTVNTQASTPOTQPOOENPP
VOATTHPEPAVTDPDLIAOPVMTVPPOPTVPRPAPPOVPOVSHPPPI
ATTPOVKTKKGVKAKADTTTPTTIPDIPHPSLAKAPAPKAKIPRESSRPKPK
DVPDSQQHPPEKSSKISEQKCCSGILKEMKAPASIVMPKPKVDVALGHDYC
DI1KHPMDSTIKSKLESREYDAQEGAVRMLFNSCYXNPNDEHVVMARKLDQV
FENRFAPMDEPEPVVTVSSPAPPTKAKKESDSSDSSDSSDSSDSSDSSDSEER
AOLAELOELKAVHQLALSOPOKPKKKEKDKKKKKKKKKKKKKKKKKKKKKKK
KELPKKTKNNSSNSVSKKEPKVPKPTPTYESSEKCKKMSYEERKRLSDIN
KLGEKIGRVVHTIIGREPLKSNPDEIIDEFTPLSTLRELERVVTGCKLKKRP
QAEKVDVLAGSSKMKGFSSSESTSSSSDSESTMAPSKKKAGHGRDOKKH
HHNHPQOPAPAPVPOPPPPPOPPPPPOPPPPPOPPPPPOPPPPPOPPPPPOPP
PPEPTAQVLEPOLGVSVDPISTFOTPLHPDELPHNLPQPEHSTPHLNH
AVSPRALHNALPOQSPRSNRALPPEKTRPAPVAPALAPOLPQMAOPPOVL
LEDEEPAPPLTSMQOLYLQLOKQVPTPLPSVVOQOPRPLPPPHSVQOOQ
LOQOPPPPPPOPPPOQOQPPPPVPHPSMPRSALIOQPPPPPOPPPPPOPP
PPPOPAKPOQVIOHPSPRHNSDPSAGHLRABPLWTHSQMFOQSLTHQSP
QOVQPKQVKGRAEPQPGVWGQGGCCEPSPAAPVPLMSQELRPSVVOQPLVV
KEEKIHSPDIRSEPFSTSLPEPEKHNENKAPVHLRQPEMKVDIGRIVRPEOS
APPKPADKDKQKQEPKVPAPKKDLKKNMSWASLVQKHPTTSSSTAKSSPSFH
FRARERKERERKAKAQAHAKEKEKEKEKEKEKEKEKEKEKEKEKEKEKEKEKE
QOOOQOOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQ
MAATIDNPFOSDLUSLIFEENLF"
BASE COUNT      1624 a      1835 c      1380 g      1161 t
ORIGIN
Alignment Scores:
Pred. No.:      3,336-132      Length:      6000
Score:          3436.50      Matches:      658
Percent Similarity: 96.66%      Conservative: 7
Best Local Similarity: 95.64%      Mismatches: 22
Query Match:     94.44%      Indels:      3
DB:              10      Gaps:      1
US-09-700-590A-22 (1-688) x AF641395 (1-6000)
QY      1 MetSerIaGluSerGlyProGlyThiArgLeuArgAsnLeuProValMetGlyAspGly 20
DB      242 ATGCTACGGAGAGCGCCCTGGACAAGATTGACAATCTGCCAGTAATGGGGATGA 301
QY      21 LeuGluThrSerGlnMetSerThrThrglnaGlnaGlnProGlnProAlaAsnAla 40
DB      302 CTAGAAACCTCCCAATATGTCACAGCAGCGCCCAACCCCAAGCCAGCAATATCA 361
QY      41 AlaSerThrAsnProProProGluThrSerAsnProAsnLeuProLysArgGlnThr 60
DB      362 GCCACACCAATCTCCACCCCGAGAGACTTCAACCTTCAACAAACCCCAAGAGACAGCA 421
QY      61 AsnGlnLeuGlnThyLeuLeuArgValValLeuLysThrLeuTrpLysHisGlnPheAla 80
DB      422 AACCACTGCATATCTGCTCAGAGTGCTGCTCAAGACACTATGAAACACCAAGTTTGG 481
QY      81 TrpProPheGlnGlnProValaAspAlaValLysLeuAsnLeuProAspTrpTrpLysIle 100
DB      482 TGGCCTTTCAGCAGACCGGTGATCCCGTCAAGCTGAACCTCCGTGATTAATAAGATT 541
QY      101 IleLysThrProMetAspMetGlyThrIleLysLysArgLeuGlnAspAsnTrpTrp 120
DB      542 ATTAAACACCAATGATGATAAGAAACAATTAAGAGCGCTTGAAACCAACTATTAATCG 601
QY      121 AsnAlaGlnGluCysIleGlnaAspPheAsnThrMetPheThrAsnCySerTrpIleTrpAsn 140
DB      602 AATGCTCAGGAAATATCTCAGAGACTTCAACACTATGTTTACAAATTTGTTACATCTATAAC 661
QY      141 LysProGlyAspAspIleValLeuMetAlaGlnaIleGlnLysLeuPheLeuGlnLys 160

```

```

DB      662 AAGCTTGAGATGACATCGCTTAAATGGCAGAAAGCTCTGAGAAAGCTCTTTCGCAAAA 721
QY      161 IleAsnGluLeuProThrgluGluThrgluIleMetIleValGlnaLysGlyArgGly 180
DB      722 ATCAATGAATCGCTTACAGAAAGAACTGAGATCATGATATGACAGGCAAAAGAAAGAGA 781
QY      181 ArgGlyValGlyGluThrglyThraAlaLysProGlyValaSerThrValProAsnThrThr 200
DB      782 CGAGGGGAGAAAGAAAGAGGGGAGCAAGAACCTGGGTATACCAAGTATCAACAAACAACT 841
QY      201 GlnAlaSerThrProProGlnThrglnThrProGlnProAsn---ProProProValGln 219
DB      842 CAGACATCAACTCTTCGCAAGACCCAGACCCCTTAGAGAAACCTCTCAACTGTGACAG 901
QY      220 AlaThrProHisProPheProAlaValaThrProAspLeuIleValGlnThrProValMet 239
DB      902 GCCCAACTACACCTTCTCTGCTGTACCCAGACCTTATGGCCAGCGCTCGTGTATG 961
QY      240 ThrValValProProGlnProLeuGlnThrProProProValaProProGlnProGlnPro 259
DB      962 ACAATGGTGCCCTCAGCCACTTCAAGACTCTTCAACCGGTACCCCCCAGCAACACCC 1021
QY      260 ProProAlaProAlaProGlnProValGlnSerHisProProIleIleAlaThrPro 279
DB      1022 CCACCTCTCTCAGTTCCACAGCTGTGCAGAGTACCCGCCCATATTCGACACACCCC 1081
QY      280 GlnProValLysThrLysLysGlyValaLysArgLysAlaAspThrThrThrThrThr 299
DB      1082 CAGCTGTGAAGAAAGAAAGAAAGGGGTGAAGAGGAACAGATACCAACACCCCTACACC 1141
QY      300 IleAspProIleHisGluProProProSerLeuProProGluProLysThrThrLysLeuGly 319
DB      1142 ATCAACCCCACTTCATGAGCCACCTCTCACTGCCCCCAAGGCCCAAGCCCAACTGGGT 1201
QY      320 GlnArgArgGluSerSerArgProValLysProProLysLysAspValaProAspSerGln 339
DB      1202 CCTGGGGGAGAGACACAGACCTGTGAAGCTCTCAAGAAAGATGTACCGGACTCACAG 1261
QY      340 GlnHisProAlaProGluLysSerSerLysValSerGlnLeuLysCysCysSerGly 359
DB      1262 CAGCACCCAGGGCCAGAGAAAGAGCAGCAGCATCTAGCAGGCTAAAGTGTGAGTGGC 1321
QY      360 IleLeuLysGluMetPheAlaLysLysHisAlaAlaTrpAlaTrpProPheTrpLysPro 379
DB      1322 ATCTCAAGAGAGATGTTGGCCAAAGAAACATGCTGCTGCTGCTCTTCAAAACCT 1381
QY      380 ValAspValaGlnaLysGluLysHisAspTrpCysAspIleIleLysHisProMetAsp 399
DB      1382 GTGATGTGGAGGCACTGGGTCTGACAGACTGTGATACATCAACAAACATCCCATGGAC 1441
QY      400 MetSerThrIleLysSerLysLeuGlnaLysArgLysLysTrpArgAspAlaGlnGluPheGly 419
DB      1442 ATGAGCAACAATCAAGTCTTAACCTAGACTCCGAGAGATGAGAGATGGCCAGAAATTTGGT 1501
QY      420 AlaAspValaArgLeuMetPheSerAsnCySerTrpLysTrpAsnProProAspHisGlnVala 439
DB      1502 GCTGATGTCCGATGATGTTCTTCCAACTGTCAACAAGTACAAACCCCTGACCAAGAAAGTG 1561
QY      440 ValaLysMetAlaArgLysLeuGlnaAspValaPheGluMetArgPheAlaLysMetProAsp 459
DB      1562 GTAGCCATAGGCTCAAAAATCTCAGAGATGTTTGAATAGCGCTTTGGCAAGATCCCTGAT 1621
QY      460 GluProGluGluProValaValaValaSerSerProAlaValaProProProThrLysVala 479
DB      1622 GAGCTGAAGAGCCAGTTTGTACAGTGTCTCTCTGCAAGTGCACCCCTACAAAGGTG 1681
QY      480 ValaLysProProSerSerSerAspSerSerSerSerSerSerSerSerSerSerSerSer 499
DB      1682 GTAGCCCAACCTCATCTAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1741
QY      500 ThrAspAspSerGluGluGluArgAlaGlnArgLeuAlaGluLeuGlnGlnLysLys 519
DB      1742 ACTGACGACTCTGAGAGAGAGAGCCAGCCAGCGGCTGACTGAATCTCAGGAACAGCTCAAG 1801

```


QY 520 A1AValHISGLIUGLNUA1A1A1eUSeGlnProGlnGlnAnLysProLysLys 539
 DB 1802 GCCGTGATGAGCAGCTTGACGCCCTCTCAAGCCAGCAAGAAACCAAGAAAAAG 1861
 QY 540 G1UlyAspLysLysG1UlyLysLysG1UlyHISLysAspLysG1UlyValG1Uly 559
 DB 1862 GAGAAAGCAAGAAAGAAAGAAAGAAAGCAAAAGAAAGAAAGAGAGAGAA 1921
 QY 560 AsnLysLysLysLysLysLysG1UlyProProLysLysLysLysAsnSerSer 579
 DB 1922 AATTAATAAAAGCAAAACCAAGAACTTCTCCAAAAGCAAGAAAGAAATACAGCAGC 1991
 QY 580 AsnSerAnValSerLysLysG1UlyProLysLysLysLysSerProProProThrTyr 599
 DB 1982 AACAGCAATGTGACAGAAAGAAAGCAAGTACCCAGAAAGCAAGCCGCTCCACATAT 2041
 QY 600 G1USeG1UlyG1UlyAspLysLysLysProMetSerLysLysLysLysLysLysLys 619
 DB 2042 GAATCAGAAAG 2101
 QY 620 LeuAspLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 639
 DB 2102 CTGATATTCACAAACTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2161
 QY 640 ArgG1UlyProSerLysLysLysLysLysLysLysLysLysLysLysLysLysLys 659
 DB 2162 AGGAAACATCATCTTAATAAATCTCCAAACCCGATGAGATTGAGATTGAGAGAGAG 2221
 QY 660 LysProSerThrLeuAAG1UlyLysLysLysLysLysLysLysLysLysLysLys 679
 DB 2222 AAGCACTCTACCTACAG 2280
 QY 680 G1UThrPheLysLeuArgLysLeu 687
 DB 2281 GAAACCTC-AAGCTGAAAAAGTTG 2303

RESULT 6
 LOCUS BC008532 2199 bp mRNA linear ROD 07-AUG-2002
 DEFINITION Mus musculus, clone IMAGE:3489640, mRNA.
 ACCESSION BC008532
 VERSION BC008532.1 GI:14250219
 KEYWORDS
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 2199)
 Strausberg, R.
 Direct Submission
 Submitted (25-MAY-2001) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Lohar Hemmighausen Ph.D., Robin Humphreys
 CDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome
 Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www-ehg.stanford.edu>
 Contact: (Dickson, Mark) med@paxi1.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
 R. M.

REMARK
 COMMENT

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
 Series: IRAX Plate: 7 Row: b Column: 12.
 This clone was selected for full length sequencing because it
 passed the following selection criteria: Hexamer frequency ORF

FEATURES
 source analysis, Similarity but not identity to protein.
 Location/Qualifiers
 1..2199
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /map="C57BL/6J"
 /clone="IMAGE:3489640"
 /tissue_type="mammary tumor. WAP-TGF alpha model. 7 months
 old, gross tissue."
 /clone_1lb="NCI CGAP_Mams"
 /lab_host="DH10B"
 /note="vector: pCMV-SPORT6"
 BASE COUNT 611 a 642 c 542 g 404 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 1,22e-96 length: 2199
 Score: 2556.50 Matches: 481
 Percent Similarity: 96.63% Conservative: 6
 Best Local Similarity: 95.44% Mismatches: 16
 Query Match: 70.25% Indels: 1
 DB: 10 Gaps: 1
 US-09-700-590A-22 (1-688) x BC008532 (1-2199)
 QY 48 ProG1UthSerAnProAnLysProLysArgG1NthAsnG1NleuG1NtyrLeuLeu 67
 DB 685 CAGAGAGCCTCCAAACCTTAACAAGCCCAAGAGACAGACAAACCAATGCAATCTGCTC 744
 QY 68 ArgValValLeuLysThrLeuTpblyshLysGlnPheAlaTrpProheGlnGlnProVal 87
 DB 745 AGAGTGCTGCTCAAGACCTATGGAACACACAGTTGCGTTCAGAGCCCGG 804
 QY 88 AspAlaValLysLeuAnLysLeuProAspTyrTyrLysLysLysLysLysLysLys 107
 DB 805 GATGCCGTCAAGCTGAACCTCTCTGATTAATTAATTAATTAATTAATTAATTAATTA 864
 QY 108 G1YThrLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 127
 DB 865 GGAACATTAAGAAGAGCGCTTGAGAAACAACTTAATGGAATGCTCAGGATGTATCCAG 924
 QY 128 AspPheAnThrMetPheThrAsnCytyrLysLysLysLysLysLysLysLysLys 147
 DB 925 GACTTAACACCTATGTTTCAAAATGTTACATTAATTAACAGCTGAGATGACATCGTC 984
 QY 148 LeuMetAlaGlnAlaLeuG1UlyLysLysLysLysLysLysLysLysLysLysLys 167
 DB 985 TTATGGCAGAGAGCTTGAGAGAGCTCTTCTGCAAAAATCAATGATGCTCTACAGAA 1044
 QY 168 G1UthG1UlyLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 187
 DB 1045 GAAACATGATATGATATGATATGATATGATATGATATGATATGATATGATATGAT 1104
 QY 188 ThrAlaLysProG1UlyLysLysLysLysLysLysLysLysLysLysLysLysLys 207
 DB 1105 GCAGCAAAAGCTGTATGATCCAGGTACCAACACAACTCAAGCATCACTCTCCGAG 1164
 QY 208 ThrGlnThrProGlnProAnLysLysLysLysLysLysLysLysLysLysLysLys 226
 DB 1165 ACCGAGAGCTCTGAG 1224
 QY 227 A1AValThrProAspLeuLysLysLysLysLysLysLysLysLysLysLysLysLys 246
 DB 1225 GCGTGAACCCAG 1284
 QY 247 LeuGlnThrProProProValProProGlnProProGlnProProProProProPro 266
 DB 1285 CTTCAGACTCTTCAACCGGTACCCCGCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1344
 QY 267 ProValGlnSerHisProProLysLysLysLysLysLysLysLysLysLysLysLysLys 286
 DB 1345 CCGTGCAAGTCAACCCGCTCATTTGAGACCAACCCCGCTGTGAGAGCAAAAGAA 1404

Oy	47	P	Pro	Glu	Thr	Ser	Asp	Pro	Asp	Asp	Pro	Leu	Arg	Gln	Ile	Thr	Ser	Asn	Gln	Ileu	Gln	Ile	Thr	Leu	Pro	66		
Db	230	C	CA	C	T	G	A	G	G	T	C	C	A	C	C	C	C	C	C	C	C	C	C	C	C	289		
Oy	67	L	eu	A	r	G	a	V	a	L	e	u	S	e	T	h	r	L	e	u	T	r	L	e	u	I	Pro	86
Db	290	C	A	G	A	T	G	T	A	G	T	G	T	A	G	A	C	T	G	T	G	A	A	C	T	G	349	
Oy	87	V	a	L	e	u	A	s	p	a	L	a	V	a	L	e	u	S	e	T	h	r	L	e	u	I	Pro	106
Db	350	G	T	G	A	T	G	A	A	T	C	A	G	A	C	T	G	A	A	T	C	A	T	A	A	A	409	
Oy	107	M	e	t	G	L	y	T	r	L	e	u	S	e	T	h	r	L	e	u	S	e	T	h	r	L	Pro	126
Db	410	A	T	G	G	A	C	T	T	A	C	A	A	G	A	C	T	T	A	C	A	A	G	A	C	T	469	
Oy	127	G	l	n	a	s	p	P	h	e	a	n	t	r	M	e	L	p	h	T	h	A	s	C	y	T	146	
Db	470	C	A	G	A	C	T	T	A	C	A	C	T	T	T	A	C	A	A	C	T	T	A	T	T	A	529	
Oy	147	V	a	L	e	u	M	e	T	a	L	a	L	e	u	G	L	y	S	e	u	P	h	e	u	G	166	
Db	530	G	T	G	T	A	T	G	C	C	C	C	C	T	T	A	G	A	A	A	T	C	T	T	C	T	589	
Oy	167	G	L	I	u	T	r	G	L	I	e	M	e	T	L	e	V	a	L	G	L	a	L	y	A	r	182	
Db	590	G	A	G	A	G	T	T	G	A	A	C	T	T	C	C	C	C	C	T	C	C	A	A	A	G	649	
Oy	183	A	r	G	L	y	S	e	u	T	r	G	L	I	r	A	L	a	L	y	P	r	O	G	L	y	202	
Db	650	G	C	C	A	A	A	T	G	A	G	T	T	C	C	A	A	A	---	G	T	G	C	C	A	A	703	
Oy	203	S	e	T	h	r	P	r	O	G	I	n	T	r	G	I	n	T	r	P	r	O	A	s	n	---	221	
Db	704	---	---	---	---	---	---	---	---	A	C	C	C	C	T	C	A	A	C	A	A	C	A	T	A	C	742	
Oy	222	P	r	O	I	S	P	r	O	P	h	e	r	O	A	L	A	T	r	P	r	O	A	S	P	L	e	241
Db	743	C	C	C	---	---	---	---	---	G	T	C	A	T	G	C	C	C	C	C	C	C	C	T	T	A	775	
Oy	242	V	a	L	P	r	O	G	I	n	P	r	O	L	e	G	I	n	T	r	P	r	O	P	r	O	V	261
Db	776	A	C	T	G	C	A	A	C	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	826		
Oy	262	A	L	A	P	r	O	A	L	P	r	O	G	I	n	S	e	r	N	i	S	P	r	O	I	L	e	281
Db	827	---	---	---	---	---	---	---	---	C	T	G	C	G	A	C	C	C	A	C	C	A	T	G	T	C	868	
Oy	282	V	a	L	y	S	t	r	L	y	S	G	L	y	V	a	L	y	A	L	a	S	P	r	T	h	r	298
Db	869	G	T	A	G	C	A	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	928	
Oy	299	---	---	---	---																							

[illegible]

AUTHORS Shang, E. and Wolgemuth, D.J.
 TITLE Cloning and expression pattern of Fsr2, a putative murine
 bromodomain-containing homolog of the Drosophila gene female
 sterile homeotic
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 2933)
 AUTHORS Shang, E. and Wolgemuth, D.J.
 DIRECT SUBMISSION
 JOURNAL Submitted (19-MAY-2000) Department of Genetics and Development,
 Columbia University, 630 W 168th St. BB1613, New York, NY 10032,
 USA

FEATURES

source

gene
CDS

Location/Qualifiers
 1..2933
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /sex="male"
 /tissue_type="testis"
 1..2933
 /gene="Fsr2"
 141..2321
 /note="similar to Drosophila FSH"
 /codon_start=1
 /product="bromodomain-containing FSH-like protein FSR2"
 /protein_id="AAF78072.1"
 /db_xref="GI:8575733"
 /translation="MSTAAAPGTPGAVPVPVPPPEVSPKPKTKTQLOQMV
 LVTLMKHAPMPFQVDAIKLNPYHGIKNPMDMGIKRLBNYWSASECQ
 DPTMTNCTYKPPDDIYLMALAEKIFLOAQMPQEVLEPPAKGRKPA
 GAQNASQOQVAVSSVSPATPONTPTVSOTVIATPTITAVTSPVPPAP
 PRPAPPIVPPPTPPVKKGVKAKDITPTTATASRSPPSLSKAKVVA
 RRESGRPIKPKKDLDEVPQAGKGLSHLRCDLSLEMLSKHAAVAMPY
 KPVDAALRLDHDIIKHPMDLSTVKKMDSREVPDAQFADIRLMFSNCKYNP
 DHEVDMARKLOVFEKRPKMPDEMEAPALPAPYAPLYSKAESRSSSESSD
 SSSSEERATRLAELOLKAHBEOLDAQAVNPKKKKKKKKKKKKKKKX
 KEKRAKSEBEKAKAPAKKQAKKAPKANKSTTASRLKKKKKKKKKKKK
 EEEGLPMSYDEKQSLDINRLPEKLGKRVVHITQREBELRNSPDEITDELK
 TTRLEIRVYKSCLOKQKRPSTSGKKQAKSKVLAQKKLELRDLQVSGQNS
 KKPTKKKSGSAPSGPSRLSSSSSSSSSSSSSSSSSSSSSSDSE"

BASE COUNT 799 a 806 c 818 g 510 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 7,7e-71 Length: 2933
 Score: 1928.00 Matches: 414
 Percent Similarity: 68.07% Conservative: 70
 Best Local Similarity: 58.23% Mismatches: 159
 Query Match: 52.98% Indels: 69
 Gaps: 18

US-09-700-590a-22 (1-688) x AF269193 (1-2933)

QY 2 SerTlaGtSerGlyPro-----GlyThrArgLeuArgAsnLeuPro 15
 DB |||||
 QY 45 AGTCCCTGATTGGCCCGAATGAGATGCCAGCCTCCAGAGAGCGCTTTTGAGC 104
 DB |||||
 QY 16 ValMetGlyAspGly-----LeuGluThrSerGlnMetSerThrThGlnAla 33
 DB |||||
 QY 105 TGCCTCTGTGACAGACAGTACAGTCCAGAGAGATCTCCTACACGCGCTCCCC 164
 DB |||||
 QY 34 GlnProGlnProAlaAsnAlaAlaSerThrAsnProProProGluThrSerAsnPro 53
 DB |||||
 QY 165 ACGGGGATCCCGGAGTCCCGGGCCCGTGAACTCTCCCACTGAGAGTCCCAACCC 224
 DB |||||
 QY 54 AsnLysProLysValGlnThrAsnGlnLeuGlnTyrLeuLeuArgValValLeuLysThr 73
 DB |||||
 QY 225 AGCAAGCCCGCGGCAAGACCTAACCACTGACATGACAGAAATGAGTGAAGACA 284
 DB |||||
 QY 74 LeuThrLysHisGlnPheAlaThrProPheGlnGlnProValAspAlaValLysLeuAsn 93
 DB |||||
 QY 285 CTCTGGAACATCACTGTTGCTTCTTCTACAGCTGTGAGATGACATCAAGCTGAAC 344
 DB |||||
 QY 94 LeuProAspTyrTyrTyrLeuIleIleLysThrProMetAspMetGlyThrIleLysLysArg 113
 DB |||||

DB |||||
 QY 345 CTCCCGATTATATATAATATATAAACCATATGATGGGACTATCAAGAAAGCA 404
 DB |||||
 QY 114 LeuGluAsnAsnTyrTyrTrpAsnAlaGlnGluCysIleGlnAspPheAsnThrMetPhe 133
 DB |||||
 QY 405 CTAGAAAATATATTATTATGAGATGCGAGTGAAGTATGACAGACTTCAACACCATGTTT 464
 DB |||||
 QY 134 ThrAsnCysTyrIleTyrAsnLysProGlyAspAspIleValLeuMetAlaGluAlaLeu 153
 DB |||||
 QY 465 ACAACGTATATATTATATAATAGCCACAGAGATATAGTGAATAGGCCCTTCA 524
 DB |||||
 QY 154 GluLysLeuPheLeuGlnLysIleAsnGluLeuProThrGluGluThrGluIleMetIle 173
 DB |||||
 QY 525 GAGAAAATCTTCTGCAAGAAAGTGCCCAAGATGCTCAGAGAGAACTGAACATATTGCC 584
 DB |||||
 QY 174 ValGlnAlaLysGlyArgGlyArg-----GlyArgGlyGluThrGlyThrAla 189
 DB |||||
 QY 585 CTGCTCCAAAGGCAAGGCGGAGGCCAGTGCAGAGAGCCCAAAATGACGTTCCCA 644
 DB |||||
 QY 190 LysProGlyValSerThrValProAsnThrThrGlnAlaSerThrProProGlnThrGln 209
 DB |||||
 QY 645 CAA-----GTGCGAGCTGTCTCTCGTCTCCAGCA----- 677
 DB |||||
 QY 210 ThrProGlnProAsn---ProProProValGlnAlaThrProHisProPheProAlaVal 228
 DB |||||
 QY 678 ACCCTCTCCAGAACATATCCCCCAACGCTGCCAAGACCC----- 719
 DB |||||
 QY 229 ThrProAspLeuIleValGlnThrProValMetThrValValProProGlnProLeuGln 248
 DB |||||
 QY 720 -----GTCAATTGCGCGTACCCCGTACCAACATCACT----- 752
 DB |||||
 QY 249 ThrProProProValProProGlnProGlnProProProProAlaProAlaProGlnProVal 268
 DB |||||
 QY 753 -----GCAACGTCAAGTCAAGTCCAGTCCCAAGCCCTCCACCGCTCTCTCGCG 806
 DB |||||
 QY 269 GlnSerHisProProIleIleAlaAlaThrProGlnProValLysThrLysGlyVal 288
 DB |||||
 QY 807 ACACCATCGCTCTGCTGCTGCTCTCCACACCG---CTGTAGTCAAGAAAAGGGCTG 863
 DB |||||
 QY 289 LysArgLysAlaAspThrThrThrProThr-----ThrIleAspProIleHis 304
 DB |||||
 QY 864 AAGCGGAAGACAGACACACACACACCAACATCTGCATCTGCACGCGAGTAG 923
 DB |||||
 QY 305 GluProProSerLeuProProGluProLysThrThrLysLeuGlyGlnAlaGlyGlnSer 324
 DB |||||
 QY 924 TCTCCCTCCACCACT---TCAGAGCCCAAGCAAGCAAGGATAGTCCCAAGGAGAGC 980
 DB |||||
 QY 325 ---SerArgProValLysProProLysValAspValProAspSerGlnGlnHisProAla 343
 DB |||||
 QY 981 GGGGCGCGCCCATCAACCTCCCAAGAGAGAGCTGGAAGATGTGAGTCCACAGAC 1040
 DB |||||
 QY 344 ProGluLysSerSerLysValSerGluGlnLeuLysCysCysSerGlyIleLeuLysGlu 363
 DB |||||
 QY 1041 GCGGCTAAAGAAAGAAACGTCTCCGACACCTGCGGACGTGACACAGCTCTCCGGAG 1100
 DB |||||
 QY 364 MetPheAlaLysLysHisAlaAlaLysAlaThrProPheTyrLysProValAspValGlu 383
 DB |||||
 QY 1101 ATGCTGTCAAGAGACATGCTGCTATGCGTGGCCCTTTTACAGCCAGGAGCGGAG 1160
 DB |||||
 QY 384 AlaLeuGlyLeuHisAspTyrCysAspIleIleLysHisProMetAspMetSerThrIle 403
 DB |||||
 QY 1161 GCGCTGAGCTGCTGCTACCATGACATCATCAAGACCCCATGACCTTCAGACGCTC 1220
 DB |||||
 QY 404 LysSerLysLeuGluAlaArgGluTyrArgAspAlaGlnGluPheGlyAlaAspValArg 423
 DB |||||
 QY 1221 AAAAGAAAGATGATACCGAGAGTACCCAGATGCAAGGGCTTTCGCTGATATCCGG 1280
 DB |||||
 QY 424 LeuMetPheSerAsnCysTyrLysTyrAsnProProAspHisGluValAlaMetAla 443
 DB |||||
 QY 1281 TTAATGTTCTCGAATTGTATTAAGTACAACCTCCACAGACCAAGAAAGTGCGCATGGCC 1340
 DB |||||
 QY 444 ArgLysLeuGlnAspValPheGluMetArgPheAlaLysMetProAspGluProGluGlu 463
 DB |||||

```

Db 1341 AGAAGCTCCAGATGTTGTGAGATGAGGTTGGCCAGATGCCCATGAGCCCATGAG 1400
Qy 464 ProValValAlaValSerSerProAlaValProProThryValValAlaProPro 483
Db 1401 -----GGGCTGGGCTGCCGCTCCAGGCGCCCATCGAGAGCAAG 1442
Qy 484 SerSerSerSerSerSerSerSerSerSerSerSerSerSerSerThraSerSer 503
Db 1443 GGGGCTGAGAGCGCCGAGTAGTGAGAGAGCTCTTCAATTGAGGACGCTCAGACTCA 1502
Qy 504 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 523
Db 1503 GAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1562
Qy 524 GlnLeuAlaAlaLeuSerGlnProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 542
Db 1563 CAGCTGGATGCTTGTCTCTCAGGCGCCAGTGAACCAAGAGAGAGAGAGAGAGAG 1622
Qy 543 LysLysGluLysLys-----LysGluLysLysLysLysLysLysLysLysLys 555
Db 1623 GAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAGCC 1682
Qy 556 GluValGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 571
Db 1683 AAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1742
Qy 572 Lys-----ThryLysLysLysLysLysLysLysLysLysLysLysLysLysLys 589
Db 1743 AAGCTCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1796
Qy 590 PrometLysSerLysProProProThryThryGluSerGluGluGluGluGluGlu 609
Db 1797 AAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1856
Qy 610 MetSerThryGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 629
Db 1857 ATGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1916
Qy 630 LeuGluArgValValValValValValValValValValValValValValValVal 649
Db 1917 CTGAGGCGGTGTGTGATCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 1976
Qy 650 AspGluLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 669
Db 1977 GACGAGATTGAGATTGAGATTGAGATTGAGATTGAGATTGAGATTGAGATTGAGATT 2035
Qy 670 LeuCysHisLeuLeuPheAlaGluGluGluGluGluGluGluGluGluGluGlu 680
Db 2036 ATATGTCAAGCTTTGTTACAAAAAAGCAGAG 2068

RESULT 9
HUMORFX
LOCUS HUMORFX 3028 bp mRNA linear PRI 06-OCT-2001
DEFINITION Human mRNA for KIAA0043 gene, complete cds.
ACCESSION D26362
VERSION D26362.1 GI:452518
KEYWORDS KIAA0043.
SOURCE Homo sapiens male myeloblast cell_line KG-1 cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 Nomura, N., Nagase, T., Miyajima, N., Saruka, T., Tanaka, A., Sato, S.,
Seki, N., Kawarabayashi, Y., Ishikawa, K. and Tsubata, S.,
Prediction of the coding sequences of unidentified human genes. II.
The coding sequences of 40 new genes (KIAA0041-KIAA0080) deduced by
analysis of cDNA clones from human cell line KG-1
DNA Res. 1 (5), 223-229 (1994)
JOURNAL MEDLINE
REFERENCE 96051398
AUTHORS 2 (bases 1 to 3028)
TITLE Ohara, O., Nagase, T., Kikuno, R. and Nomura, N.
AUTHORS Direct Submission
JOURNAL Submitted (27-DEC-1993) Osamu Ohara, Kazusa DNA Research Institute;

```

```

FEATURES
source
1532-3, Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail:cdna1info@kazusa.or.jp, Tel:+81-438-52-3913)
Location/Qualifiers
1..3028
/organism="Homo sapiens"
/db_xref="taxon:9606"
/sex="male"
/cell_line="KG-1"
/cell_type="myeloblast"
1..3028
/gene="KIAA0043"
1..139
/gene="KIAA0043"
5' UTR
140..2320
/gene="KIAA0043"
CDS
140..2320
/note="similar to Human homolog of Drosophila female
sterile homeotic mRNA ( HUMSHG )"
/codon_start=1
/protein_id="BAA05393.1"
/db_xref="GI:452519"
/translation="MSTATVAPAGIPATPGVPNPPEVNSPKRGRKTNOLOYMON
VVKTLMKHGFAMPEYOPVDALKINLPYHKIKRPMGMGTIKRLENNYVWASCEM
ODFNTMPTNCYIYNKPTDDIVLMAQALEKIQKAVKQPOBEVELLPAPKGRKRA
AGAQSAGTQVVAANSVSPATPQSPPTVSOTPVIAATPVTTIANVTSVPVPAVA
PEPATPIVPVPPPTPVKKGVRKADTTPTTSATITASESPPLSDPKQAVV
ARRBSGGRPIKPKKDLDEGEVPOHAGKGLSEHLRYCDILREMLSKHAAVMP
KRPVDALELDYDIIKHPMDLSTVRKMDGREYPAQGPADVRLMFSCTYKYP
PDHEVVAARLQDVFEMRFAPKMPDEYEPALPAPAPVMSKGAESSRSSSESDS
GSSDSEERATRLAEIQOLKAVHEQALASQAPNKRKKKKKKKKKKKKKKKKKK
EKHKVAKEEERKAKVAPPAKQAOQKAAKAKNSTTGTGRQURKGGKASASDSEE
EGLPMSYDERKQLSLDINRLPGLKGVVHIIQREPSLRNSDDEIIEPTLKPT
TLRELERYKSCLOKQKRPFSAGSKQAKSKEBLAQEKLEKRLQDVSGQLSS
KKPARKRGSPSGGSPRLSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS
2321..3028
/gene="KIAA0043"
3' UTR
BASE COUNT 758 a 864 c 877 g 529 t
ORIGIN
Alignment Scores:
Pred. No.: 1,39e-70 Length: 3028
Score: 1922.00 Matches: 403
Percent Similarity: 69.51% Conservative: 69
Best Local Similarity: 59.35% Mismatches: 151
Query Match: 52.82% Indels: 57
DB: Gaps: 16

US-09-700-590A-22 (1-688) x HUMORFX (1-3028)
Qy 25 GlnMetSerThryGlnAlaGlnAlaGlnProGlnProAlaAsnAlaAlaSerThr 43
Db 137 AGGATGTCACCGCCGACGACGACGTCGCGCGGGGATCCCGGACCCCGGCCCTGTG 196
Qy 44 AsnProProProProGluThrySerAnProAnLysProLysArgGlnThraGlnLeu 63
Db 197 AACCCACCCCCCGGAGGCTCTCCAAACCCACAGAGCCGCGGCGCAAGCCAAACGAGCTG 256
Qy 64 GlnThryLeuLeuArgValValLeuLysThryLysHisGlnPheAlaTrpProPhe 83
Db 257 CAGTACATCAGAAATGTGTGTGTAAGAGCGCTGTGAAACACACAGATTCCCTGGCCCTTC 316
Qy 84 GlnGlnProValAspAlaValLysLeuAnLeuProAspTyrThryLysLysLysLysThr 103
Db 317 TACCAAGCCCGTGAAGCCATCAAAATTTGAACCTGCGGATTAATCAATAATTATTAATAAC 376
Qy 104 PrometLysPmeGlyThryLysLysArgLeuGluAsnAsnTyrTyrTrpAsnAlaGln 123
Db 377 CCAATGATATGGGCGACTATTAAAGAAAGAACTAGAAAATTAATTAATTAATTAATTAATTA 436
Qy 124 GluCysLysLysLysPheLysThrMetPheThraGlnCysTyrLysLysLysProGly 143
Db 437 GAATGTATGACGAGACTTCAACCAACATGTTTACAAATTTTATTAACAAGCCACACA 496
Qy 144 AspAspIleValLeuMetAlaGluAlaLeuGluLysLeuPheLeuGlnLysIleAsnGlu 163

```

```

Db 497 GATGACATAGTGCTAATGAGCCCAAGCTTTAGAGAAATTTTCTACAAAAGTGGCCAG 556
Qy 164 LeuProthrlugluThrlgullileMetlleValGlnAlaIysGlyArgGlyArg----- 181
Db 557 ATGCCCCAAGAGGAAAGTTATTTATCCCTGCTCCAAAGGGGCAAGGCTGGAAACCC 616
Qy 182 -----GlyArgIysGluThrlgullileMetlleValGlnAlaIysGlyArgGlyArg----- 199
Db 617 GCGCGGAGACCCAGAGCGGAGTACACAGCA-----GTGGGGCGCGTGTCTCTGTC 670
Qy 200 ThrGlnAlaSerThrProProGlnThrlgullileMetlleValGlnAlaIysGlyArgGlyArg----- 218
Db 671 TCCCGACCG-----ACCCCTTTTCAGACCGTGGCCCGCCACCGCTC 709
Qy 219 GlnAlaThrProHisProPheProAlaValThrProAspLeuIleValGlnThrlgullileMetlleValGlnAlaIysGlyArgGlyArg----- 238
Db 710 TCCGACGAGCC-----GTACATCGCTGCCACCGCTGTA 742
Qy 239 MetThrValValProProGlnProLeuGlnThrlgullileMetlleValGlnAlaIysGlyArgGlyArg----- 258
Db 743 CCAACCATCATCTGAAC-----GTACATCGCTGCCACCGCTGTA 793
Qy 259 ProProAlaProAlaProGlnProAlaGlnSerHisProProIleIleAlaIleThr 278
Db 794 CCACTCTCT-----CTGCGCACCCATCGTCCCGTGTCTCTCTACAG 838
Qy 279 ProGlnProAlaIysThrIysGlyValIysArgIysAlaAspThrThrThrProThr 298
Db 839 CCG-----CCTGCTCGTCAAGAAAGGCGCTGAAGCGGAAAGGACACCAACCATCTCCACG 895
Qy 299 -----ThrIleAspProIleHisGluProProSerLeuProProGlnProIle 314
Db 896 ACCTGCGGCATCATCTGCCAGCGGAGTGTGCGCCCGCCGCTG-----TCAGACCCCAAG 952
Qy 315 ThrThrIysLeuGlyGlnArgArgGluSer-----SerArgProValIysProProIle 333
Db 953 CAGCGCAAGTGTGTGCGCGCGGAGAGTGTGTGCGCGCCATCAAGCTCCCAAGAG 1012
Qy 334 AspValProAspSerGlnGlnHisProAlaProGluIysSerSerIysValSerGluGln 353
Db 1013 GACTGTGAGGAGCGGAGGTGCCCGACGACGACGAGCAAGAAAGGCGCAACTTGTGAGAC 1072
Qy 354 LeuIysCysCysSerGlyIleLeuIysGluMetPheAlaIysIysHisAspTyrCysAspIle 373
Db 1073 CTGGGCTACTGCGACAGATCTCTAGGAGATCTATCTCAAGAAAGCGCGCTTACGCC 1132
Qy 374 TrpProPheTyrIysProValAspValGlnAlaIysGlyLeuHisAspTyrCysAspIle 393
Db 1133 TGCCCTTCTCAAGCGAGTGTGATGCCAGGCCCTGTGAGCTGCACTACCGACATC 1192
Qy 394 IleIysHisProMetAspMetSerThrIleIysSerIysLeuGlnAlaIysGluTyrArg 413
Db 1193 ATCAAGACCGCGATGTGACTGTGACCGCTGAAAGAAAGATGAGGCGCGAGATGCCA 1252
Qy 414 AspAlaGlnGluPheGlyAlaAspValArgLeuMetPheSerAsnCysTyrIysTyrAsn 433
Db 1253 GAGCGACAGGCGCTTGTCTGTGATGTCCGCTGATGTCTGAAATTGCTTACAAATACAT 1312
Qy 434 ProProAspHisGluValAlaIleMetAlaIysGlyLeuGlnAspValPheGluMetArg 453
Db 1313 CCCCCAGACCGAGGTGTGTGCGCTGCGGAGGCTCCAGAGCTGTGTTGAGATGAGG 1372
Qy 454 PheAlaIysMetProAspGluProGluGluProValAlaIleValSerSerProAlaVal 473
Db 1373 TTTGCGCAAGATGCGAGATGAGCGC-----GTGAGGCGACCGCGCTG 1414
Qy 474 ProProProThrIysValAlaIleProProSerSerAspSerSerSerAspSerSer 493
Db 1415 CTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1474
Qy 494 SerAspSerAspSerSerThrAspAspSerGluGluArgAlaGlnArgLeuAlaGlu 513

```

```

Db 1475 AGCTCTTGCGACTCAGGAGCTCGGACTCGGAGAGAGCGGCGCCACGCTGGCGAG 1534
Qy 514 LeuGlnGluIleuIysAlaValHisGluGlnLeuAlaLeuSerGlnProGln 533
Db 1535 CTGACAGAGAGAGCTGAAGGCGGTGACAGAGAGCTGCGCGCTGTCTACAGGCCAGTA 1594
Qy 534 AsnIysProIysIys-----LysGluIysAspIysIysGluIys 547
Db 1595 AACAAACCAAGAAAGAAAGAAAGAAAGAAAGAGAGAGAGAAAGAAAGAGAGAGAG 1654
Qy 548 LysGluIysHisIysArgIysGluGlnValGlnIleAsnIysIysSerIys----- 564
Db 1655 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1714
Qy 565 ---AlaIysGluProProIysIysThrIysAsnAsnSerSerAsnVal 583
Db 1715 CTGCGCAAGAGAGCTACAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1774
Qy 584 SerIysIysGluProAlaProMetIysSerIysProProThrThrIysGluGln 603
Db 1775 GCGGAGAGAGAGCTGAAGAAAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1834
Qy 604 GluAspIysCysIysProMetSerTyrGlnIysArgIysIleuSerIleuAspIleAsn 623
Db 1835 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1894
Qy 624 LysIleuProGluIysIysLeuGlyArgValHisIleIleGlnSerArgIysProSer 643
Db 1895 CGGCTGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1954
Qy 644 LeuIysAsnSerAsnProAspGluIleGlnIleAspPheGluThrIleuIysProSerThr 663
Db 1955 CTCAGGAGACTCAACCCCGAGAGATGAATTTGACTTTGAGACTGTGAAGAACCCACACT 2014
Qy 664 LeuArgIleuGlyAlaLeuCysHisLeuIysPheAlaGluGlu---LysGluThr 681
Db 2015 TTGGGGGAACT-GGAGAGATGTCAAGTCTTTTACAGAAAGAGAGAGAGAGAGAGAG 2070

RESULT 10
AF045462 3649 bp mRNA linear ROD 01-JUL-1998
LOCUS AF045462
DEFINITION Mus musculus female sterile homeotic-related protein Frig-1 mRNA,
complete cds.
ACCESSION AF045462
VERSION AF045462.1 GI:3273700
KEYWORDS
SOURCE Mus musculus.
ORGANISM Mus musculus.
REFERENCE 1 (bases 1 to 3649)
AUTHORS Rhee, K., Brunori, M., Besset, V. and Wolgemuth, D.J.
TITLE Expression and potential role of Frig-1, a putative murine
sterile homeotic
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3649)
AUTHORS Rhee, K., Brunori, M., Besset, V. and Wolgemuth, D.J.
TITLE Direct Submission
JOURNAL Submitted (29-JAN-1998) Department of Genetics and Development,
Columbia University College of Physicians and Surgeons, 630 W 168th
St., Black Bldg 1633, New York, NY 10032, USA
FEATURES
source
1..3649
/organism="Mus musculus"
/db_xref="taxon:10090"
681..3077
/note="bromodomain-containing protein"
/codon_start=1
/product="female sterile homeotic-related protein Frig-1"
/protein_id="AAC24810.1"
/db_xref="GI:3273701"
/translation="MLQNTVPHKLPEGNAGLGLGPBAAPGRIRKPSLLYGFES

```

```

PTMASVPAIQLAPNPPEVSNPKKPGVTVNQLYHKVVKALMKHOFAMPFPROPV
DAVKGLPDYHKI IKOPMDGT I KRLENNYWAASECQADNTMTNICYINPKPD
IYMAQLEKLEI FLQKVASMOEOLVITI PKNSHKQSKLALQSLTSAHVAVS
SVSHATLYPPEIPTTVANIHPHSVTSPLKSLHSPAPLGAAPAPDAPLKKK
GVKRAADTPPTPTALAGSPASPGSLSPKRAPIPMRRESGPRVPRDLPDPSO
QOHOSKKKSEBQKHCNGIILKELSKHAAVMPFYRPVDSALGLHDYDIKH
MULSTVKRMENRDRDAQEPADVRLMNCYINPDPHDVAMARKLDQVFEERYA
KMPDELEGPVPVSTALPPGLTKSSSESSSESSSESSSESSSESSSESSSESS
DEBERAHLALQOLRAVHQAALASQPIISKPRKREKKKKKRAKFKHGRIG
IDEDKPPAPRPPPOPKSKKAGGGGNNITLSPFGSGSSNKL PKRSQKAPV
LPTGYDEEESESRPMVSDEKROLIDIKLPEKIGKRYVHI IOAREPSRNSPEI
EIDFTLKPSSTIREBERYVLSCKRKRPRPYITIRKVGKTKBELAEKRELEKLDQ
VSGQINSTKPKPKASEKTESAQLVPVRLSASSSSSDSSSSSSSSSDTSDSDG
"
BASE COUNT      920 a      986 c      976 g      767 t
ORIGIN
Alignment Scores:
Score:           1,2e-69      Length:      3649
Percent Similarity: 66.94%      Matches:      403
Best Local Similarity: 55.51%      Conservative: 83
Query Match:      52.24%      Mismatches: 145
DB:               10          Indels:      96
                       Gaps:       18
US-09-700-590A-22 (1-688) x AF045462 (1-3649)
QY      5 SerGlyProGlyThrArgLeuArgAsnLeuProValMetGlyAspGlyLeuGluThrSer 24
DB      756 GCACGACCCGAGAAAGATTCCAAAGCCTTCTGCTGTATAGAGGATTGGAGAGCC 815
QY      25 GlnMetSerThrThrglnAlaGlnAlaGlnProGlnProAlaAsnAlaAlaSerThrAsn 44
DB      816 ACAATGGCTTCTGTACAGCT-----TTACAACGCGCCCTCGCCCAAT 857
QY      45 ProProProProGluThrSerAsnProAsnLysProLysArgGlnThrAsnGlnLeuGln 64
DB      858 CCACGACCCCTGAGGTGTCCTCAACCAAAAGCCAGAGCGGGTACCAAAACCAACGTCGAG 917
QY      65 TyrLeuLeuArgValValLeuLysThrLeuTyrLysHisGlnPheAlaTyrProPheGln 84
DB      918 TACCTCAACAGGATGATGATGAGGCTCTGTGGAAGCATGATTTGCATGCGCAATTCGG 977
QY      85 GlnProValAspAlaValLysLeuAsnLeuProAspTyrTyrLysIleIleLysThrPro 104
DB      978 CAGCCTGTGACCGCTGGAAGCTGGGTTCCCGGATTATACAAATTTAAACAGCCCT 1037
QY      105 MetAspMetGlyThrIleLysLysArgLeuGluAsnAsnTyrTyrTyrPheAsnAlaGlnGlu 124
DB      1038 ATGACATGAGGTACTATCAAGAGAGACTTGAAACCAATTACTAGCGGCTGCTCAGAA 1097
QY      125 CysIleGlnAspPheAsnThrMetPheThrAsnGlyTyrIleTyrAsnLysProGlyAsp 144
DB      1098 TGTATCAGAGATTATTATATCTATGTTTACCAACTGTTTATTAACAAGCCACCGCAT 1157
QY      145 AspIleValLeuMetAlaGlnAlaLeuGluLysLeuPheLeuGlnLysIleAsnGluLeu 164
DB      1158 GATATTGTCTTAATGCGACAGACACTGGAAAAATCTTCTTAAGAAAGTGCATCATG 1217
QY      165 ProThrGluGluThrGluLeuMetIleValGlnAlaLysGly-----ArgGly--- 180
DB      1218 CCACAGAGAGGACAGAGCTTGTGTGACCATCCCTAAAAACGACCTAAGAGGGGGG 1277
QY      181 -----ArgGlyArgLysGluThrGlyThrAlaLysProGlyValIleSerThr 195
DB      1278 AAGTAGACAGCACTCCAGGACGATTAATCAAGCCCATCAGGTGCTGTCTTCTTCT 1337
QY      196 ValProAsnThrThrGlnAlaSerThrProProGluThrGlnThrProGlnProAsnPro 215
DB      1338 GTGTCCATACACCGCTGTATACACCACTCGAAATA-----CCT 1379
QY      216 ProProValGlnAlaThrProHisProPheProAlaValThrProAspLeuIleValGln 235

```

```

DB      1380 ACCACTGTCTCAACATTCGCCACCA-----TCAGTCATCTCT 1418
QY      236 ThrProValMetThrValValProProGlnProLeuGlnThrProProProValProPro 255
DB      1419 TCTCCTCTTCTTAAGTCCCTGAT-----TCTGCTGAGACCCCATCCTTGTCT 1466
QY      256 GlnProGlnProProAlaProAlaProGlnProValGlnSerHisProProIleIle 275
DB      1467 GTATCAGACGCGCTCAGCT----- 1487
QY      276 AlaAlaThrProGlnProValLysThrLysGlyValLysArgLysAlaAspThrThr 295
DB      1488 -----CAGCCCTTCGCAAGAAAAGGCGTTAAACGGAACCGGATTACTTACC 1535
QY      296 ThrProThrThrIle-----AspProIleHisGluProProSerLeu 309
DB      1536 ACCCTTAACCCACAGCACTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1595
QY      310 ProProGluProLysThrThrLysLeu---GlyGlnArgArgGluSerSerArgProVal 328
DB      1596 -----GAGCCAAAGGACAGAGAGCTCCTCTCATGCGCAGAGAGTGGCGCCCAATC 1649
QY      329 LysProProLysLysAspValProAspSerGlnGlnHisProAlaProGluLysSerSer 348
DB      1650 AAACCCCAAGAAAGACTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1709
QY      349 LysValSerGluGlnLeuLysCysCysSerGlyIleLeuLysGluMetPheAlaLysLys 368
DB      1710 AACGCTTCAGAGCATTAAGCATGCTCAACGCGATCTTAAGAACTGCTCTTAAGAAAG 1769
QY      369 HisAlaAlaTyrAlaTyrProPheTyrLysProValAspValGluAlaLeuGlyLeuHis 388
DB      1770 CAGCTGCTCAGCGCTGCGCTCTCTTAAGCAGTGAAGCTTCTGCTGCTGCTGCTGCT 1829
QY      389 AspTyrCysAspIleIleLysHisProMetAspMetSerThrIleLysSerLysLeuGlu 408
DB      1830 GATTACCATGACATCACTTAACAACCCCATGACCTCAGACGTCGTAAGGGAGAGTGGAG 1889
QY      409 AlaArgGluTyrArgAspAlaGlnGluPheGlyAlaAspValArgLeuMetPheSerAsn 428
DB      1890 AACCGTACACTCGGATGACAGAGATTGCTGCTGATGATGAGGCTTATGTTCTTCAAC 1949
QY      429 CysTyrLysTyrAsnProProAspHisGluValAlaMetAlaArgLysLeuGlnAsp 448
DB      1950 TGCTATAAGTACATCTCCACAGACGATGTTGTGCTATGCGACAGAAAGTTGCGAGAT 2009
QY      449 ValPheGluMetArgPheAlaLysMetProAspGluProGluGluProValValAlaVal 468
DB      2010 GTGTTTGAAGTTTCGCTATCCAAAGATGACAGATGAGCCACTGGAACGAGACCTTACCA 2069
QY      469 SerSerProAlaValProProPro---ThrLysValValAlaProProSerSerSerAsp 487
DB      2070 GTCTTAATGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2129
QY      488 SerSerSerSerSerSerSer-----AspSer 496
DB      2130 AGTAGAGTAGAGATTCCTCTGACGAGAGAGAGAGAGAGAGAGAGATGAGACGAGAG 2189
QY      497 AspSerSerThrAspAspSerGluGluGluArgAlaGlnArgLeuAlaGluLeuGlnGlu 516
DB      2190 GAGAGTAGAAGCTCAGACTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2249
QY      517 GlnLeuLysAlaValHisGluGlnLeuAlaAlaLeuSerGlnProGlnGlnLysPro 536
DB      2250 CAGCTTCGAGGAGTTATATACACTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2309
QY      537 LysLysLysGluLysAspLysLysGluLysLysLys-----GluLysHisLys 552
DB      2310 AAGCGGAGAG---AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2366
QY      553 ArgLysGluGluValGluGluLysLysSer---LysAlaLysGluProPro---Pro 570
DB      2367 GGCCTAATGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2426

```


Db 1778 TCTCCACTTCAAGTCTTTCAC-----TCTGCTGAGACCCCGCTCTTGTCT 1825
 QY 256 GlnProGlnProProAlaProAlaProGlnProValGlnSerHisProProIleIle 275
 Db 1826 GTTACTGACACTCTCTCAGC----- 1846
 QY 276 AlaAlaThrProGlnProValLysThrLysGlyValLysArgLysAlaAspThrThr 295
 Db 1847 -----CAGCCCTTGGCCCAAGAAAAAGGCGTAAAGCGAAGAGCATACTACC 1894
 QY 296 ThrProThrThrIle-----AspProIleHisGluProProSerLeu 309
 Db 1895 ACCCTTACACCTACAGCATCTTGGCTCTGCTTCTCCAGCTAGCCCTCTCGGAGCTT 1954
 QY 310 ProProGluProLysThrThrLysLeu---GlyGlnArgArgGlnSerSerArgProVal 328
 Db 1955 -----GAGCTTAAGGAGAGACCGCTTCCCCCTATGCGTGAAGAGAGATTGGTCCGCCATC 2008
 QY 329 LysProProLysLysAspValProAspSerGlnGlnHisProAlaProGluLysSerSer 348
 Db 2009 AACCCCCACGCAAGACTTGCCTGACTCTCAGCAACACACAGCTCTTAAGAAAGCA 2068
 QY 349 LysValSerGluGlnLeuLysCysCysSerGlyLysLeuLysGluMetPheAlaLysLys 368
 Db 2069 AACCTTTCACAAACAGTTAAACATTGCAATGGCATTTTGAAGGAGTTACTCTTAAGAG 2128
 QY 369 HisAlaAlaThrAlaThrProPheThrLysProValAspValGluAlaLeuGlyLysHis 388
 Db 2129 CAGCTGCTCCTATGCTTGGCTTCTTATTAACCACTGATGCTTCTGCACTTGGCTGCTCAT 2188
 QY 389 AspTyrCysAspIleIleLysHisProMetAspMetSerThrIleLysSerLysLeuGlu 408
 Db 2189 GACTACCATGACATCATTAAGACACCCCATGACTGACTGCAAGCGAAGATGAG 2248
 QY 409 AlaArgGluLysArgAspAlaGlnGluPheGlyAlaAspValArgLeuMetPheSerAsn 428
 Db 2249 AACCGTATTAACCGGAGTGCACAGAGATTGCTGCTGATATACCGCTTATGTTCTTCAAC 2308
 QY 429 CysTyrLysTyrAsnProProAspHisGluValAlaAlaMetAlaArgLysLeuGlnAsp 448
 Db 2309 TGCCTATTAAGTACATCCCCCAGATCAGCATGTTGTGCAATGGCAGCAAGCTACAGGAT 2368
 QY 449 ValPheGluMetArgPheAlaLysMetProAspGluProGluGluProValAlaAlaVal 468
 Db 2369 GATATTGAGTTCGCTTATGCAAGATCCAGATGAAACACAGAACACGAGGCCCTTACCA 2428
 QY 469 SerSerProAlaValProProPro---ThrLysValValAlaProProSerSerSerAsp 487
 Db 2429 GTCCTTACTCTCATGCCCCCTGGCTTGGCCAAATCGTCTTCAGAGTCTCCAGTGAAGAA 2488
 QY 488 SerSerSerSerSerSer-----Asp 495
 Db 2489 AGTAGCAGTAGAGCTCTCTGAGGAAAGAGAGAGAGATGAGAGACGAGAGAGAA 2548
 QY 496 SerAspSerSerThrAspAspSerGluGluValArgAlaGlnArgLeuAlaGluLeuGln 515
 Db 2549 GAAAGAGCTAAAGCTCAGACTCAGAGGAAAGAAAGGCTCATCGCTTAAGCAAGACTACAG 2608
 QY 516 GluGlnLeuLysAlaValHisGluGlnLeuAlaLeuSerGlnProGlnGlnAsnLys 535
 Db 2609 GAACAGCTTCGGGCGAGTACGAAACAATGCTGCTCTGCCAGGGCTCCAAATATCCAG 2668
 QY 536 ProLysLysLysGluLysAspLysLysGluLysLysLys-----GluLysHis 551
 Db 2669 CCCAAGAGGAAA---AGAGGAAAAAAGAAAGAAAGAAAGCAAGAGCAAGAGAACAT 2725
 QY 552 LysArgSerGluGluValGluGluAsnLysLysSer---LysAlaLysGluProPro--- 569
 Db 2726 CGAGGCGGAGCTGGGGCGGACGAAGATGACAGAGGCGCTTAGGACCCCGCCACCTCA 2785
 QY 570 ProLysLysThrLysLysAsnAsnSerSerAsnSerAsnValSerLysLysGluProAla 589
 Db 2786 CCTAAGAAAGTCAAGAAAGCAAGTGGGCGGTGGCAGTGGCTTGAAGGCGCTTCT 2845

QY 590 -----ProMetLysSerLys 594
 Db 2846 GGCTTGAAGCTTCTGAGAGAGTGCACCAAGCTCCCCAAAAAGGCCCAAGAGACAGCC 2905
 QY 595 ProPro-----ProThr---TyrGlnSerGluGluLysAspLysCysLysProMetSer 611
 Db 2906 CCACCTTCCCTGCTCAGATTATGATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2965
 QY 612 TyrGluGluLysArgGlnLeuSerLysLeuAspIleAsnLysLeuProGluGluLysLeuGly 631
 Db 2966 TACGATGAGAAAGCGGAGCTGAGCTGAGCATTAACAATTACTCTGGGAGAAAGCTGGC 3025
 QY 632 ArgValValHisIleIleGlnSerArgGluProSerLeuLysAsnSerAsnProAspGlu 651
 Db 3026 CGAGTTGTGATATATCAACGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3085
 QY 652 IleGluIleAspPheGluThrLeuLysProSerThrLeuArgGluLeuGlyAlaLeuCys 671
 Db 3086 ATTGAGATTGATTGTAACACTCAAGCCATCCACCTTAAGAGAGAGAGAGAGAGAG 3144
 QY 672 HisLeuLeu 674
 Db 3145 CCTTCTCTG 3153
 RESULT 12
 HUMFSHG 4053 bp mRNA linear PRI 22-JUN-1995
 LOCUS Human homolog of Drosophila female sterile homeotic gene, complete cds.
 DEFINITION
 ACCESSION M80613.1 GI:182768
 VERSION M80613.1
 KEYWORDS MHC class II antigen; female sterile homeotic gene.
 SOURCE Homo sapiens CDNA to mRNA.
 ORGANISM Homo sapiens
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 4053)
 AUTHORS Beck,S., Hansen,I.M., Kelly,A., Pappin,D.J.C. and Trowsdale,J.
 TITLE A homologue of the Drosophila female sterile homeotic (Fsh) gene in the class II region of the human MHC
 JOURNAL DNA Seq. (1991) in press
 REFERENCE 2 (sites)
 AUTHORS Haynes,S.R., Dollard,C., Winston,F., Beck,S., Trowsdale,J. and Dawd,I.B.
 TITLE The bromodomain: a conserved sequence found in human, Drosophila and yeast proteins
 JOURNAL Nucleic Acids Res. 20 (10), 2603 (1992)
 MEDLINE 92285152
 PUBMED 1350857
 FEATURES
 source location/Qualifiers
 1..4053 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /cell_type="T CELL"
 1178..3442 /note="putative"
 /codon_start=1
 /protein_id="AA68890.1"
 /db_xref="GI:182769"
 /translation="MASVPAQLQLPANPPPEVNSPKRGRVTNQLQYLHKVMKALM
 KHQFAMPPOPDAVKLGLPDYHKLIKQPMQGTIKRLNNYVAASECDPNTMF
 TNCYLTKNPETDVIDMAQTLKELIKLOKASMPQOEOLVITPKSHKGAALTAQ
 SVTSAHQVPAVSVAHTLALYTPRPITPTVNIIPRVSYSPLLSLSAGPILAVT
 NAPPQAPLAKKKGVKRAKADTTPTATLALRGSPPASPGSLRPAKLPPMRBSGRP
 IKPRKDPDSQQQKSSKGLSGLKHCNGILKELLSKRAAVAMPYKPDVASAL
 GLHDYHDIKHEMDISTVRKMEKNDYDAQFAADVRLMFSNCTKYNPDPDHYVAMA
 RKLQVFEFRVAKMPDEPLEPGLPVSTAMPGLAKSSSSSESSSESSSESEEE
 DEDEDEDESSDESEERARHARLQOLAVHQLAALSGPISKPKRREKKEK
 KKRKAEKRGKRGADDEDKGRAAPRPPPKSKSKASGSGGSAALGPGFGSGSGCT
 KLPKATVTAAPALPTGYDSESEESRPSVDEKQUSUDIKLFGELKGRVHIIQA
 REPSLRDNPEIETDFTLKPSLTRELRVLSCLRKKPRKPYITKKRVGTXKELA
 LEKKRELKRLQDVSGQLNSTKPKPKKANETESSAOOVAVSRISASSSSSSSSSS

	polyA_signal	4015..4020	SSSSSDTSDSOG"
	/note="putative"		
BASE COUNT	997 a	1132 c	1067 g 857 t
ORIGIN			
Alignment Scores:			
Pred. NO.:	4.48e-69	Length:	4053
Score:	1888.00	Matches:	404
Percent Similarity:	67.08%	Conservative:	81
Best Local Similarity:	55.88%	Mismatches:	142
Query Match:	51.88%	Indels:	97
DB:	9	Gaps:	20
US-09-700-590A-22 (1-688) x HUMFSHG (1-4053)			
Oy	5 serGlyProGIYThrArgLeuArgAsnLeuProValMetGIYAspQIYLeuGIUnThrSer	24	
Db	1115 GCAGACCCAGGGAAAGATTCCAAAACCCTCTCTGTGTAGAAGCGTTGAGAGCCCC	1174	
Oy	25 GlnMetSerThrArgInalagInalagInProGlnProAlaSnlaAlaSerThrAsn	44	
Db	1175 ACAATGGCTTCGGTGCTGCCT---TTGCACTTAACCCCTGCC-----AAC	1216	
Oy	45 ProProPProPProGIUnThrSerAsnProAsnLysProLYsArgGlnThrAsnGlnLeuGln	64	
Db	1217 CCACACACCCCGAGGCTGTCCAATCCCAAAAAACCGAGAGGTTCACAACCACTGCAA	1276	
Oy	65 TyrLeuLeuArgValValLeuLysThrLeuTrpLYsHIGlnPhelAtPProPheGln	84	
Db	1277 TACTTACCAAGTAGTAGTAGTAAGAGCCTCTGTGAACAATCAAGTCCCATGGCCATTCCGG	1336	
Oy	85 GlnProValAspAlaValLYsLeuAsnLeuProAspTYrTYrLYsILEIleLYsThrPro	104	
Db	1337 CAGCCTGGATGCTGTCCAACCTGGGCTCTCCGATTATCACAAATATAAAGCAGCCT	1396	
Oy	105 MetAspMetGIYThrLIleLYsArgLeuGlnAsnAsnTYrTYrTrpAsnlaGlnGlu	124	
Db	1397 ATGGACATGGGTACTATTAAAGAGAGACTTGAAACAAATATTATTGGGCTGCTTCAGAG	1456	
Oy	125 CysIleGlnAspPheAsnThrMetPheThrAsnCYS TYrILETYrAsnLYsProGIYAsp	144	
Db	1457 TGTAAGCAAGTTTTTAATACATGTTACCAACCTGTACTTATACAAACAAGCCACTGAT	1516	
Oy	145 AspIleValLeuMetAlaGlnAlaLeuGlnLYsLeuPheLeuGlnLYsILEasnglueu	164	
Db	1517 GATATTGCTTAATGGCAACAAGCCTGGAAAAGATATTCCTACAGAGAGGTTCATCATATG	1576	
Oy	165 ProThrGlnGluThrGluILEmetILEvalGlnAlaLYsGLY-----ArgGLY--	180	
Db	1577 CCACAMGAAGAACAAAGACTGGTAGTATGATCCATCCCTAAGAACAGCCACMAAGGGGGCCC	1636	
Oy	181 -----ArgGLYArgLYsGluThrGLYThrAlaLYsProGIYAlaSerThr	195	
Db	1637 AAGTTGGCAGGCGCTCCAGGGCAGTGTATACAGATGCCCAACAGATGCTCTCTCTCT	1636	
Oy	196 ValProAsnThrArgInalaserThrProProGlnThrGlnThrProGlnProAsnPro	215	
Db	1697 GTGTACACACAGCCCGCTGTATATCTCCACCTGAGATA-----CCT	1738	
Oy	216 ProProValGlnAlaThrProHisPProPheProAlaValThrProAspleuILEvalGln	235	
Db	1739 ACCACTGCTCCCAACATTCACCCACCCA-----TCACCTATTCC	1777	
Oy	236 ThrProValMetThrValProProGlnProLeuGlnThrProProProValProPro	255	
Db	1778 TCTCCACTTCTCAAGTCTTGAC-----TCTGCTGAGACCCCGCTCTCTGCT	1825	
Oy	256 GlnProGlnProProProAlaProAlaProGlnProValGlnSerHisProProIleile	275	
Db	1826 GTTACTGAGGCTCTCCAGCC-----	1846	
Oy	276 AlaAlaThrProGlnProValLYsThrLYsLYsGLYValLYsArgLYsAlaAsnThrThr	295	

[illegible]


```

Db 3034 GATTATGATTCCTTATGCCAGATGCCAGATGACCACTAGAACAGAGGCGCTTTACCA 3093
Qy SerSerProAlaValProProPro---ThrIysValAlaProProSerSerSerAsp 487
Db 3094 GTCTTACTGCAAGCCCCCTGGCTGGCCAAATCGCTTCAAGACTCTCCAGTGAGGAA 3153
Qy 488 SerSerSerAspSerSerSer-----Asp 495
Db 3154 AGTAGCAGTAGAGACTCCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 3213
Qy 496 SerAspSerSerThrAspAspSerGluGluGluValGluGluGluGluGluGluGlu 515
Db 3214 GAAGAGAGTAGAAGCTCAAGACTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3273
Qy GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 535
Db 3274 GAACAGCTTCGGGAGTAGACATGACAACTGGCTGCTCTGCTCCAGGCTCAATATCCAA 3333
Qy 536 ProIysIysIysGluIysAspIysIysGluIysIysIysIysIysIysIysIysIys 551
Db 3334 CCAAGAGAGAGAA---AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3390
Qy 552 LysArgIysGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 569
Db 3391 CGAGCGCGAGCTGGGGCCGAGAGATGACAAAGGGGCGCTTACAGGCCCGCCCACTCAA 3450
Qy 570 ProIysIysThrIysIysAspAsnSerSerAsnSerAsnValSerIysIysGluProAla 589
Db 3451 CCTAAGAGATCCAAAGAAAGCAAGTGCGAGTGGGGGTGGCAGTGTCTTGTAGGCCCTTCT 3510
Qy 590 -----PrometIysSerIys 594
Db 3511 GGCTTTGGACCTTCTGAGAGAGTGCCACCAAGCTCCCCAAAAGGCCAACAAAGACAGCC 3570
Qy 595 ProPro-----ProThr---TyGluSerGluGluGluGluGluGluGluGluGluGlu 611
Db 3571 CCACCTGCCTCGCTCAAGCTTATGATTCAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 3630
Qy 612 TyGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 631
Db 3631 TACAGATGAGAGAGCGGACCTGAGCTGAGCATCAACAAATTACTTGGGAGAGAGAGCTGGGC 3690
Qy 632 ArgValAlaHisIleIleGlnSerArgGluProSerIleIysAsnSerAspProAspGlu 651
Db 3691 CGAGTGTGCATATATATCAAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3750
Qy 652 IleGluIleAspPheGluThrIleIysProSerThrIleArgGluGluGluGluGluGlu 671
Db 3751 ATTGAGATTGATTTGAAACACTCAAGCCATCCACATTAGAGAGCTTGA--GCGCTATGT 3809
Qy 672 HisIleLeu 674
Db 3810 CCTTCTCTG 3818

RESULT 14
LOCUS HUMKIA9001 4664 bp mRNA linear PRI 09-FEB-1999
DEFINITION Human mRNA for KIAA9001 gene, complete cds.
ACCESSION D42040
VERSION D42040.1 GI:577292
KEYWORDS KIAA9001.
SOURCE Homo sapiens male myeloblast cell_line KG-1 cDNA to mRNA.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 4664)
AUTHORS Nomura,N.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 4664)
AUTHORS Nomura,N.
TITLE Direct Submission
JOURNAL Submitted (07-NOV-1994) Nobuo Nomura, Kazusa DNA Research
Institute, Gene Structure 1; 1532-3 Yana, Kisarazu, Chiba 292,

```

```

FEATURES
    source          Japan (E-mail:cdnainfo@kazusa.or.jp, URL:http://www.kazusa.or.jp,
                    Tel:0438-52-3930, Fax:0438-52-3931)
                    Location/Qualifiers
                        1..4664
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /sex="male"
                        /cell_line="KG-1"
                        /cell_type="myeloblast"
                        1..4664
                        /gene="KIAA9001"
                        5' UTR
                        1..1701
                        /gene="KIAA9001"
                        CDS
                        1702..4107
                        /gene="KIAA9001"
                        /note="female sterile homeotic (fsh) homolog RING3"
                        /codon_start=1
                        /protein_id="BA07641.1"
                        /db_xref="GI:577293"
                        /translation="MLONTVPHNKLDEGNAGLIGLPEAAAPGRKIRKPSLLYGEF
                        /SPMASVPAIOLTPANPPEPVSNPKPKRVTNQLYLHKVVKALMKIOFAMFROP
                        VDAVKGLPDYHKIIRKQMDMGIRKLENNYMAASECQODENTFTCYIYNKPTD
                        DIYMAQTLTEKIFLOKVASMPQEBELVVTIPKNSHGKAKLAALQSYSAQVAV
                        SVSHTAITYTPPEITPTLYNIPHSVTSIPILKSHSGAPLITATAPPAQPLAKK
                        KGVKRAVDITTPPTATILAGSPASPGLSEPAALPPMKRESGRPIPRPDLPDS
                        QOQHOSKKKLESEQLKNGILKELLSKHAAYAPFYKPVASALGLDHYDITKH
                        PMLOSTVKKRMENRDYDQEPADVRLMFSNCKYKPNPDHVDVAARKLOVFEERY
                        AKMDEPLRPGPLPSTAMPPGIAGSSSESSSESSSESESESESESESESESESESESE
                        SPSEREERARHLAEIOELRAYEOLALISQGPISKPKRREKREKKRKAKHGR
                        AGADEDDKPRARPRPPPKSKSAAGSGGSAIAGSGGSGGSGGTKLPKYATAP
                        PALPTGVDSSESESESRPMSIDEKQSLDINKIPKRLGRVHIIQARPSLDSNP
                        EIIDETLKPSTLRELYVLSQLRKRPKPYTIKPVGKTEBIALEKRELRKL
                        ODVSGQLNSTTKPKPKKANETBSSAQOVAVSRLSASSSSSSSSSSSSSSSDTS
                        DSG"
BASE COUNT      1084 a 1336 c 1293 g 951 t
ORIGIN
Alignment Scores:
Score:           5,11e-69          Length: 4664
Percent Similarity: 1888.00        Matches: 404
Best Local Similarity: 67.08%      Conservative: 81
Query Match:     55.88%            Mismatches: 142
DB:              51.88%            Indels: 97
                       Gaps: 20
US-09-700-590A-22 (1-688) x HUMKIA9001 (1-4664)
Qy 5 SerGluProGluThrArgLeuArgAsnLeuProValMetGlyAspGlyLeuGluThrSer 24
Db 1780 GCAGACACGAGGAGAAAGATTCGAAACCCCTCTCTGTATGAGGCGCTTGAAGAGCCC 1839
Qy 25 GlnMetSerThrThrGlnAlaGlnAlaGlnProGlnProAlaAsnAlaIleSerThrAsn 44
Db 1840 ACAATGGCTTCGGTGGCTGCTGCT---TTGCACACTTACCCCTGCG-----AAC 1881
Qy 45 ProProProProGluThrSerAspProAsnIysProIysArgGlnThrAsnGluGln 64
Db 1882 CCACCACCCCGGAGAGGTCCAAATCCAAAAGCCAGAGAGAGAGAGAGAGAGAGAGAGAG 1941
Qy 65 TyrlleuArgValIleLeuIysThrIleuTrpIysHisGlnPheAlaTrpProPheGln 84
Db 1942 TACCTACACAGTAGTAGATGAAGGCTCTGTGGAAACATCATGTTCCGATGGCCATTCGG 2001
Qy 85 GlnProValAspAlaValIysLeuAsnLeuProAspTyrTyrIysIleIleIysThrPro 104
Db 2002 CAGCCTGTGATGCTGTCAAACTGGGTCTTACCGGATTAATCAAAAATTAATAAACAGCCT 2061
Qy 105 MetAspMetGlyThrIleIysLeuArgLeuGluGluGluGluGluGluGluGluGluGlu 124
Db 2062 ATGACATGGGTACTATTTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2121

```

QY 125 CyeIIeglnaspheasnrhmetptheThAsnCysfryrlleryrAsnlyspProGlyAsp 144
 DB 2122 TGTATGCAAGATTTTATATACCATGTTTACCAACGTTTATACCAAGCCACATGAT 2181
 QY 145 AspIIeValleuwerAlaGluAlaLeuGluLeuPheLeuGlnlyslleAsnGluLeu 164
 DB 2182 GATATTGTCCTAATGGCAAAAGCCTGAAAAGATATTCTTACAGAAAGTTGCATCATG 2241
 QY 165 ProThrgluGluThrgluIleMetIleValGlnAlaIysGly-----ArgGly--- 180
 DB 2242 CCACAGAAAGAAACAGAGCGTAGTACCATCCCTTAAGAAACGCCACAGAAAGGGGGCC 2301
 QY 181 -----ArgGlyArgGlyGluThrglyThralaIysProGlyIysSerThr 195
 DB 2302 AAGTTGGACAGCGCTCCAGGCGAGTGTACCAAGTCCATGAGGCGCTGCGCTCTTCT 2361
 QY 196 ValProasnrhThrglnAlaSerThrProProGlnThrglnThrProGlnProasnrPro 215
 DB 2362 GTGTCAACACAGCCCTGTATACTCTTCCACCTGAGATTA-----CCT 2403
 QY 216 ProProValGlnAlaThrProHisProPheProAlaValThrProAspLeuIleValGln 235
 DB 2404 ACCACTGTCTCAACATCTCCACCA-----TCAGTCATTTC 2442
 QY 236 ThProValMetThrValValProProGlnProLeuGlnThrProProProValProPro 235
 DB 2443 TCTCACTCTCAAGCTTCGAC-----TCTGCTGAGCCCGCTCTTGTCT 2490
 QY 256 GlnProGlnProProProAlaProAlaProGlnProValGlnSerHisProProIleIle 275
 DB 2491 GTTACTGACAGCTCTCCAGC----- 2511
 QY 276 AlaAlaThrProGlnProValIysThrIysGlyValIysArgIysAlaAspThrThr 295
 DB 2512 -----CAGCCCTTGGCCAAAGAAAAAGCGCTTAAGCGAAAGAGATACTACC 2559
 QY 296 ThProThrThrIle-----AspProIleHisGluProProSerLeu 309
 DB 2560 ACCCTACACCTACAGCATCTTGGCTCTGTTCTCTCCAGTACCTCTCGGAGTCTT 2619
 QY 310 ProProGluProIysThrThrIysLeu--GlyGlnArgArgIysSerArgProVal 328
 DB 2620 -----GAGCCTTAAGCGACAGCGGCTTCCCTATGCGTGAAGAGTGTGCGCCATC 2673
 QY 329 LysProProIysIysAspValProAspSerGlnGlnHisProAlaProGluIysSerSer 348
 DB 2674 AACCCCCACGCAAGACCTTGCTGACTCTGACCAACACCAAGAGCTCTTAAGAAAGGA 2733
 QY 349 LysValSerGluGlnLeuIysCysCysSerGlyIleLeuIysGluMetPheAlaIysLys 368
 DB 2734 AACCTTCAAGAACAGTTAAACATTCGATGGCATTTTGAAGAGATTACTCTTAAGAG 2793
 QY 369 HisAlaAlaIysAlaTrpPheThrIysProValAspValGluAlaLeuGlyLeuHis 388
 DB 2794 CAGTGTGCTCATGTGGCTTCTTATAAACAGATGATCTTCTGCACTTGGCTGAT 2853
 QY 389 AspDlyCysAspIleIleIysHisProMetAspMetSerThrIleIysSerIysLeuGlu 408
 DB 2854 GACTACCATATCATTAAGACCCCATGACTGACCTGCACTGCAAGCGGAAGATGAG 2913
 QY 409 AlaArgGluIysArgAspAlaGlnGluPheGlyAlaAspValArgLeuMetPheSerAsn 428
 DB 2914 AACCGTATTAACGGGATGACAGAGTTTGGCTGATGATGATGATGATGATGATGATGAT 2973
 QY 429 CyeTyrIysfryrAsnProProAspHisGluValAlaAlaMetAlaArgIysLeuGlnAsp 448
 DB 2974 TGTATTAAGTACATCCCCAGATTCAGATGTTGTGCAATGGACCAAGGCTACAGAGAT 3033
 QY 449 ValPheGluMetArgPheAlaIysMetProAspGluProGluGluProValAlaAlaVal 468
 DB 3034 GATATTGAGTTCCGTATGCGCAAGATCCAGATGAACCATAGAACAGGAGCCCTTATCA 3093
 QY 469 SerSerProAlaValProProPro--ThrIysValValAlaProProSerSerSerAsp 487

DB 3094 GTCTTACTGCTCCATGCCCCCTGCTTGCCAAATCGTCTTACAGATCTTCACTGAGGAA 3153
 QY 488 SerSerSerSerSerSer-----Asp 495
 DB 3154 AGTAGAGTACAGAGCTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 3213
 QY 496 SerAspSerSerThraspAspSerGluGluIysArgAlaGlnArgLeuAlaGluLeuGln 515
 DB 3214 GAAGAGGTGAAGAGCTCAGACTCAGAGAGAAAGAGGCTCATGCTTAGAGAACTACAG 3273
 QY 516 GluGlnLeuIysAlaValHisGluGlnLeuAlaIleLeuSerGlnProGlnGlnAsnLys 535
 DB 3274 GAACAGCTTCGGGACAGTACATGAACAATGCGCTGCTGTCCAGAGGTCCAAATATCAAG 3333
 QY 536 ProIysIysGluGluIysAspIysGluIysLysLys-----GluIysHis 551
 DB 3334 CCCAAGAGGAAA--AGAGGAAAAAGAGAAAAAGAAAGAACGGAAGGACAGAGAGCAT 3390
 QY 552 LysArgIysGluGluValGluIysLeuIysSer--LysAlaIysGluProPro-- 569
 DB 3391 CGAGGCCAGCTGGGGCCGATGAAGATGAACAAGGGCTTAAGGCCACCCGCCACCTCAA 3450
 QY 570 ProIysIysThrIysIysAsnAsnSerSerAsnSerAsnValSerIysGluProAla 589
 DB 3451 CTTAAGAGTCCAAAGAAACAGAGTGGAGTGGGCTGAGAGTCTCTTACGCTCTTCT 3510
 QY 590 -----PrometIysSerLys 594
 DB 3511 GGCTTTGACCTTCTGAGAGAGTGGACCAACGCTCCCAAAAAAGCCACAAAGACAGCC 3570
 QY 595 ProPro-----ProThr--TyrGlnSerGluGluIysAspIysCysLysProMetSer 611
 DB 3571 CCACCTGCCCTGCTCAGATTAATGATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3630
 QY 612 TyrGluGluIysArgIysLeuSerIysLeuAspIleAsnIysLeuProGluIysLeuGly 631
 DB 3631 TACGATGAGAGAGCGCAGCTGAGCTGACATCAACAAATTACTCTGGGAGAGAGCTGGG 3690
 QY 632 ArgValValHisIleIleGlnSerArgGluProSerLeuIysAsnSerAsnProAspGlu 651
 DB 3691 CGAGTTGTGATTAATCCAGAGCGAGAGCCCTCTTTCAGTATTCAAAACCAAGAGAG 3750
 QY 652 IleGluIleAspPheGluThrLeuIysProSerThrLeuArgIysGluLeuAlaLeuCys 671
 DB 3751 ATTGAAGATTGATTTTAAACACTCAAGCCATCCACACTTAGAGAGCTTGA--GGCTATGT 3809
 QY 672 HisLeuLeu 674
 DB 3810 CTTTCCTG 3818
 RESULT 15
 BC032124 2548 bp mRNA linear PRI 26-JUN-2002
 LOCUS Homo sapiens, Similar to bromodomain containing 3, clone MGC:29612
 DEFINITION IMAGE:485640, mRNA, complete cds.
 ACCESSION BC032124
 VERSION BC032124.1 GI:21594669
 KEYWORDS MGC.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 2548)
 AUTHORS Strausberg, R.
 TITLE Direct Submission
 JOURNAL Submitted (06-JUN-2002) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 REMARK NIH-MGC Project URL: <http://mgc.ncbi.nih.gov>
 CONTACT Contact: MGC help desk
 COMMENT Email: cgabbs-r@mail.nih.gov

Tissue Procurement: ATCC
 CDNA Library Preparation: Rubin Laboratory
 DNA Sequencing by: The I.M.A.G.E. Consortium (LNLN)
 DNA Sequencing by: National Institutes of Health Intramural
 Sequencing Center (NISC),
 Gaithersburg, Maryland;
 Web site: <http://www.nisc.nih.gov/>
 Contact: nisc.mc@nih.gov

Akhter, N., Ayle, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
 Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
 Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Hashighi, P.,
 Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Latic, P., Legaspi, R.,
 Maduro, Q.L., Mastello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,
 McDowell, J., Pearson, R., Stancirpop, S., Thomas, P.J., Touchman, J.W.,
 Tsougen, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
 Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LNLN at: <http://Image.lnl.gov>
 Series: IRAL Plate: 39 Row: 9 Column: 22

This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 12408642.

FEATURES

Source

1. 2548
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="MGC:29612 IMAGE:4856840"
 /tissue_type="Colon, adenocarcinoma"
 /clone_lib="NIH MGC_15"
 /lab_host="DH10B-R"
 /note="Vector: pOTB"
 308..1978
 /codon_start=1
 /product="Similar to bromodomain containing 3"
 /protein_id="AAH32124.1"
 /db_xref="GI:21594670"
 /translation="MSTATTVAAPGIPATTEGPVNPPEVSNSEKPERKTNQLOLYON
 VVVTILKHOFPAMFPYQVDALILNLPDHYKILNPDMGIIKRLNNYWSSECM
 QDFMTWNTCYIYNKPRDILVMAQALEKIFLQVAPQDEVLLEPPAKGKRKA
 AGASACTOOAAVSVPATFQSVPTVQPTVATPVPTTAVTAVSVPPA
 PPPATPTVPPPTPPVKKYKGRADPTTPTSAITSRSSPPSLDPPQAKV
 ARBSGGRPIKPKKLEDEGVPOHAKGKLSHLKXCSILREMLSKGAAYWPF
 YKPDVALLELHDHDIIRKPMDSITVKKRKGDEYDAQFADVLNFSCKYNP
 PKHEVVMARKLDQFEMRPAKMPDEVEPALPAPAPVNSKESRSSESSDS
 GSSDEERARLAELOELQALVAHQALASQAVNPKKKKEKKEKKKDEKEXK
 EKHKVKAEEKKAKVAPAKQAOQKAPAKKANSTTTAGRDHFLTGV"

CDS

BASE COUNT 641 a 754 c 753 g 400 t
 ORIGIN

Alignment Scores:

Pred. No.: 7,82e-69 Length: 2548
 Score: 1877.50 Matches: 408
 Percent Similarity: 61.33% Conservatave: 63
 Best Local Similarity: 53.12% Mismatches: 152
 Query Match: 51.59% Indels: 146
 Gaps: 18

US-09-700-590a-22 (1-688) x BC032124 (1-2548)

QY 25 GlnMetSerThrThrGlnAlaGlnAlaGlnProGln---ProAlaAsnAlaAlaSerThr 43
 Db 305 AGGATGTGTCACCGCACGACAGTCGCGCGGGAGATCCCGCGACACCGCGGCGCTGTG 364
 QY 44 AsnProProProGlnThrSerAsnProAsnLysProLysArgGlnThrAsnGlnLeu 63
 Db 365 AACCCACCCCGCGAGGCTCCCAACCCCGAGCCGCGCGGACCAACAGCTG 424
 QY 64 GlnTyrLeuLeuArgValValLeuLysThrLeuTyrLysHisGlnPheAlaTyrProPhe 83
 Db 425 CAGTTCATGCAAGATGTGTGTGAAGAGCTCTGGAACAACAAGTTCGCGCGCTTC 484
 QY 84 GlnGlnProValAspAlaValValLysLeuAsnLeuProAspTyrTyrLysIleIleLysThr 103

Db 485 TACGAGCCCGTGCAGCGCAATCATCAATTTGAACCTCGGATTTATCATTAATTTAAAC 544
 QY 104 ProMetAspMetGlyThrIleLysLysArgLeuGlnAsnAsnTyrTyrPasnAlaGln 123
 Db 545 CCAATGATATGGGAGACTATTAAAGAGAGACTAGAAAATATATTATTATGTGAGGACGC 604
 QY 124 GluCysIleGlnAspPheAsnThrMetPheThrAsnCysTyrIleTyrAsnLysProGly 143
 Db 605 GAATGTATGAGAGCTTCAACACCACTGTTTACAAATGTTTACATTATTAACAAGCCACA 664
 QY 144 AspAspIleValLeuMetAlaGlnAlaLeuGlnLysLeuPheLeuGlnLysIleAsnGln 163
 Db 665 GATACCATATGCTCAATGAGCCCAAGCTTTAGAGAAATTTTCTACAAAAAGTGCCACG 724
 QY 164 LeuProThrGlnGlnThrGlnIleMetIleValGlnAlaLysGlyArgGlyArg----- 181
 Db 725 ATGCCCAAGAGAAAGTTGATATTATTCCTCCCTCCCAAGGGCAAGGTCGGACCGC 784
 QY 182 -----GlyArgLysGlnThrGlyThrAlaLysProGlyValSerThrValProAsnThr 199
 Db 785 GCTGCGGAGAGCCGACGAGCGAGTACACAGCAA-----GTGGCGCGCTGTCTGTCTC 838
 QY 200 ThrGlnAlaSerThrProProGlnThrGlnThrPro---GlnProAsnProProProVal 218
 Db 839 TCCCCACCG-----ACCCCTTTCAGAGCGTCCGCCACCGCTC 877
 QY 219 GlnAlaThrProHisProPheProAlaValThrProAspLeuIleValGlnThrProVal 238
 Db 878 TCCAGACGCGC-----GTCACTCGCTGCACCCCTGTA 910
 QY 239 MetThrValValProProGlnProLeuGlnThrProProProValProProGlnProGln 258
 Db 911 CCAACCATCACTGCAAC-----GTCACTCGCTGCACCCCTGTA 961
 QY 259 ProProProAlaProAlaProGlnProValGlnSerHisProProIleIleAlaAlaThr 278
 Db 962 CCACCTCT-----CTGCCACACCCATCGTCCCGGAGTCCCTCAAG 1006
 QY 279 ProGlnProValLysThrLysLysGlyValLysArgLysAlaAspThrThrProThr 298
 Db 1007 CCG---CCTGTCGCAAGAAAAGGCGTGAAGGGAAGGAGACACACACACTCCACG 1063
 QY 299 -----ThrIleAspProIleHisGlnProProSerLeuProProGlnProLys 314
 Db 1064 ACCTGCGCATCATCTGCACGCGGAGAGTGAAGTCCGCGCGCTG---TCAGACCCCAAG 1120
 QY 315 ThrThrLysLeuGlyLysArgArgLysSer---SerArgProValLysProProLysLys 333
 Db 1121 CAGGCAAGTGTGTGCGCGCGGAGAGTGTGAGCGCCGCTCCATCAAGCTCCCAAGAG 1180
 QY 334 AspValProAspSerGlnGlnHisProAlaProGlnLysSerSerLysValSerGlnGln 353
 Db 1181 GACCTGAGAGAGCGGCGAGGTGCGCCGAGAGTGTGAGCGCCGCTCCATCAAGCTCCCAAGAG 1240
 QY 354 LeuLysCysCysSerGlyIleLeuLysGlnMetPheAlaLysLysHisAlaAlaTyrAla 373
 Db 1241 CTGCGCTACTGCGACAGATCTCTCAGGAGATGCTATCCAAAGACAGCGCGCTACGCGC 1300
 QY 374 ThrProPheTyrLysProValAspValGlnAlaLeuGlnLysHisAspTyrCysAspIle 393
 Db 1301 TGGCCCTTTCACAGCCAGTGTGATGCGGAGCGCTTGAGCTGCACAGTACAGACATC 1360
 QY 394 IleLysHisProMetAspMetSerThrIleLysSerLysLeuGlnAlaArgGlnTyrArg 413
 Db 1361 ATCAAGACCCGAGTGAAGCTCAGACCGTGAAGAAAGAGATGAGCGGAGATACCA 1420
 QY 414 AspAlaGlnGlnPheGlnAlaAspValArgLeuMetPheSerAsnCysTyrLysTyrAsn 433
 Db 1421 GACGACAGGCGCTTGTGCTGATGTCGCGGTGATGTTCTGAATTGCTGAATAATACAT 1480
 QY 434 ProProAspHisGlnValAlaValAlaMetAlaArgLysLeuGlnAspValPheGlnMetArg 453
 Db 1481 CCCCAGACCAAGAGTTGTGCGATGCGCCGGAAGCTCCAGAGCGTGTGAATGAGG 1540

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 14, 2003, 18:42:55 ; Search time 369 Seconds
(without alignments)
4198.848 Million cell updates/sec

Title: US-09-700-590a-22

Perfect score: 3639
Sequence: 1 MSASGPGTRLRNLPVMDG.....ALCHLLFAEKEKTFKRLKM 688

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 1125999159 residues
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame_p2n.model -DEV=xlp
-O=/cgn2.1/USPTO-spool/US9700590/runat.14032003.140629.3214/app.query.fasta_1.839
-DB=N_Geneseq_101002 -QFMT=fastlap -SUFFIX=eng -MINMATCH=50.1 -LOOPEXT=0
-LOOPEXT=0 -UNITS=bits -STAR=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -NORM=exc -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USBR=US9700590 @CGN 1.1 396 @runat.14032003.140629.3214 -NCPU=6 -ICPU=3
-NO_XLPPX -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -IONLOG -DEV TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_101002.*

1: /SID3/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SID3/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SID3/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SID3/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SID3/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SID3/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SID3/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SID3/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SID3/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SID3/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SID3/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SID3/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SID3/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SID3/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SID3/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
16: /SID3/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SID3/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SID3/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SID3/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SID3/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SID3/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SID3/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SID3/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SID3/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	3639	100.0	2329	AAZ56719	Human transmembran
2	1888	51.9	4664	ABK83660	Human CDNA differe
3	1888	51.9	4664	ABN95159	Gene #1657 used to
4	1501.5	41.3	3104	AAV68343	Transcriptional re
5	1501.5	41.3	3106	AAK40172	WO9904265 Seq ID N
6	1450	39.8	1884	AAH18591	Human CDNA sequenc
7	1377	37.8	6776	ABL04337	Drosophila melanog
8	1237.5	34.0	1741	ABK84162	Human CDNA differe
9	1161.5	31.9	1862	ABV25740	Human prostate exp
10	981	27.0	667	AA503029	Human diagnostic a
11	688	18.9	746	AA526193	Human CDNA encodin
12	564	15.5	424	AA526606	Human CDNA encodin
13	484.5	13.3	450	AAZ80432	Human colon cancer
14	457	12.6	560	ABQ60244	Human polynucleotide RTP
15	430.5	11.8	642	AAK29128	Drosophila melanog
16	430	11.8	1542	ABL23049	Drosophila melanog
17	430	11.8	3542	ABL23048	Drosophila melanog
18	411.5	11.3	559	AAA43730	Mouse secreted exp
19	391	10.7	352	AAK66021	Novel human polynu
20	344	9.5	2247	AAK69456	Arabidopsis thalia
21	323.5	8.9	7326	AAV10093	Sequence encoding
22	323.5	8.9	7326	AAV10093	Cellular transcrip
23	323.5	8.9	7326	AAK62948	CREB binding prote
24	323.5	8.9	7326	AAK62948	Mouse nuclear CREB
25	318.5	8.8	807	ABL19177	Drosophila melanog
26	318.5	8.8	2807	ABL19176	Drosophila melanog
27	318.5	8.8	5017	ABL15790	Drosophila melanog
28	308.5	8.5	547	AA557537	cDNA #213 encoding
29	306	8.4	2442	AAK50976	Arabidopsis thalia
30	305	8.4	9046	AAK07792	Transcription fact
31	305	8.4	9046	AAV10092	Cellular transcrip
32	305	8.4	9046	ABL49723	Human p300 gene SE
33	303	8.3	597	ABV57572	Human prostate exp
34	302	8.3	209	AAV87340	EST clone BM2. Ho
35	302	8.3	881	AAK56273	Pinus radiata tran
36	300.5	8.3	6743	AAK65172	DNA encoding novel
37	300.5	8.3	6743	AAK65172	DNA encoding novel
38	300.5	8.3	6743	AAK65172	Drosophila melanog
39	282	7.7	1445	AAK47126	Arabidopsis thalia
40	281.5	7.7	3593	ABL09563	Drosophila melanog
41	281.5	7.7	6161	ABL09562	Drosophila melanog
42	277.5	7.6	2014	AAK45013	Arabidopsis thalia
43	275.5	7.6	30690	AAA92301	S. avermitilis ave
44	275.5	7.6	30690	AAH79277	Streptomyces averm
45	274.	7.5	14648	ABL16120	Drosophila melanog

ALIGNMENTS

RESULT 1
AAZ56719 standard; CDNA; 2329 BP.

AAZ56719;
23-MAR-2000 (first entry)

DE Human transmembrane protein HTPMPN-22 encoding CDNA.
XX
XX Human; transmembrane protein; HTPMPN; diagnosis; immunospecific;
KW antiproliferative; neuroprotective; immune disorder;
KW reproductive disorder; smooth muscle disorder; neurological disorder;
KW gastrointestinal disorder; developmental disorder;
KW cell proliferative disorder; ss.
XX
XX Homo sapiens.
XX

PN MO961471-A2.
 XX
 PD 02-DEC-1999.
 XX
 XX 28-MAY-1999; 99MO-US11904.
 XX
 XX 29-MAY-1998; 98US-0087260.
 PR 02-JUL-1998; 98US-0091674.
 PR 02-OCT-1998; 98US-0102954.
 PR 24-NOV-1998; 98US-0109869.
 XX
 XX (INCY-) INCYTE PHARM INC.
 XX
 PI Tang YT, Lai P, Hillman JL, Yue H, Guegler KJ, Corley NC;
 PI Bandman O, Patterson C, Gorgone GA, Kaser MR, Baughn MR;
 PI Au-Young J;
 XX
 DR WPI; 2000-072605/06.
 DR P-PSDB; AAY57898.
 XX
 PT Proteins, polynucleotides, vectors, host cells and antibodies used to
 PT diagnose, treat or prevent immune, reproductive, smooth muscle,
 PT neurological, gastrointestinal, developmental and cell proliferative
 PT disorders -
 XX
 XX
 XX Claim 9; Page 191; 229pp; English.
 XX
 XX
 CC AAZ56698 to AAZ56776 encode AAY57877 to AAY57955 which represent human
 CC transmembrane proteins designated HTMPN-1 to HTMPN-79, respectively.
 CC The transmembrane protein have immunospecific, antiproliferative and
 CC neuroprotective activities. The human transmembrane proteins,
 CC polynucleotides encoding them and other compositions and methods from
 CC the present invention, can be used for the diagnosis, treatment or
 CC prevention of immune, reproductive, smooth muscle, neurological,
 CC gastrointestinal, developmental and cell proliferative disorders. The
 CC HTMPN's can be used to treat or prevent disorders associated with a
 CC decreased expression or activity of HTMPN.
 XX
 SO Sequence 2329 BP; 674 A; 691 C; 569 G; 382 T; 13 other;
 Alignment Scores:
 Pred. No: 1.09e-166 Length: 2329
 Score: 3639.00 Matches: 688
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 21 Gaps: 0
 US-09-700-590a-22 (1-688) x AAZ56719 (1-2329)
 QY 1 MetSerIaGluSerGlyProGlyThrArgLeuArgAsnLeuProValMetGlyAspGly 20
 DB 16 ATGCTGGGAGAGGCGCCCTGGGACGAGATTGAGAAATCTGCCATATGGGGATGGA 75
 QY 21 LeuGluThrSerGlnMetSerThrThrglnAglAglAglProGlnProAlaAsnAla 40
 DB 76 CTAAAACTTCCCAATGCTTCAACACACAGGCCGACCAACCCCAAGCCCAACCA 135
 QY 41 AlaserThrAnpProProProGluThrSerAnpProAsnLysProLysArgGlnThr 60
 DB 136 GCCAGCAACCAACCCGCCCCCAAGACCTTCAACCTTCAACCAACCAAGGAGGAGCC 195
 QY 61 AsnGlnLeuGlnIleuLeuArgValIleuLysThrLeuTrpLysHisGlnPheAla 80
 DB 196 AACCAACGCAATACCTCTCAAGATGGTGCACAGACCTATGGAACACCAAGTTTCA 255
 QY 81 TrpProPheGlnGlnProValAspAlaValLysLeuAsnLeuProAspTyrTyrLysIle 100
 DB 256 TGGCTTTCCAGCAGCCTCTGATGCCGTCAAGCTCAAGCTCTCCGATTAATAGATC 315
 QY 101 IleLysThrProMetAspMetGlyThrIleLysLysArgLeuGluAsnAsnTyrTyrTrp 120
 DB 316 ATTAAGCGCTATGATATGGAACTAAAGAGCGCTTGGAAGAAACAATATTACTG 375

QY 121 AsnAlaGlnGluCysIleGlnAspPheAsnThrMetPheThrAsnGlyTyrIleTyrAsn 140
 DB 376 AATGCTGAGAAATGATATCCAGCACTTCAACACTATGTTTCAATATGTTACTTACAC 435
 QY 141 LysProGlnLysAspAspIleValLeuMetAlaGlnAlaLeuGluLysLeuPheLeuGlnLys 160
 DB 436 AAGCTTGAGATGACATAGCTTTAATGGCAGAACTCTTGAAAGGCTCTTTCGCAAAA 495
 QY 161 IleAsnGluLeuProThrThrGluGluThrGluIleMetIleValGlnAlaLysGlyArgGly 180
 DB 496 ATAAATAGCTTACCCACAGAAAGAACCCAGATCATGATATGTCAGGCAAAAGAGAGGA 555
 QY 181 ArgGlyArgLysGluThrGlyThrAlaLysProGlyValSerThrValProAsnThrThr 200
 DB 556 CGTGAGAGAAAGAAACAGGAGACAGCAAACTGGCGTTTCCACGGTATCAACACAACT 615
 QY 201 GlnAlaserThrProProGlnThrGlnThrProGlnProAsnProProProValGlnAla 220
 DB 616 CAAGCATCGACTCTCCCGCAGACCACAGACCCCTCAGCCGAATCTCTCTGTGACAGGC 675
 QY 221 ThrProHisProPheProAlaValThrProAspLeuIleValGlnThrProValMetThr 240
 DB 676 AGCCTCAACCTTCTCTGCGTCAACCCGAGCTCATCTCCAGACCTCTGTATACA 735
 QY 241 ValValProProGlnProLeuGlnThrProProValProProGlnProGlnProPro 260
 DB 736 GTGTGCTCTCCCGCAGCAGACAGCCCGCCGAGTGCCCGCCCGCAGCAACCCCA 795
 QY 261 ProAlaProAlaProGlnProValGlnSerHisProProIleIleAlaIleThrProGln 280
 DB 796 CCCCCTCAGTCCCGCCCGCCGTAACAAGCACCACCACTATCTGGGGCACCACCA 855
 QY 281 ProValLysThrLysLysGlyValLysArgLysAlaAspThrThrThrProThrThrIle 300
 DB 856 CCTGTGAAGAAAGAGAGAGAGTGAAGAGAAACACACACACCCCAACCACTT 915
 QY 301 AspProIleHisGluProProSerLeuProProGluProLysThrThrLysLeuGlyGln 320
 DB 916 GACCCATTCAAGACGACACCTCGTCCCGGAGCCCAAGACCAACAGCTGGGACAG 975
 QY 321 ArgArgLysSerSerArgProValLysProProLysLysAspValProAspSerGlnGln 340
 DB 976 CGGCGGAGAGACACCCGCTGTGAAACCTTCAAAAGAGACGCGCCGCTCTCACAG 1035
 QY 341 HisProAlaProGluLysSerSerLeuValSerGluGlnLeuLysCysCysSerGlyIle 360
 DB 1036 CACCAGACCAAGAGAGAGACGACAGGCTCTGGAGCACTCAAGTCTGACGGGACATC 1095
 QY 361 LeuLysGluMetPheAlaLysLysHisAlaAlaTyrAlaTrpProPheTyrLysProVal 380
 DB 1096 CTCAAGAGAGATGTTTGCAGAAAGACACCGCCCTAGCCCTTGCCCTTCAACACTGTG 1155
 QY 381 AspValGluAlaLeuGlyLeuHisAspTyrCysAspIleIleLysHisProMetAspMet 400
 DB 1156 GACGTGAGGACACTGGGCTCAACACTACTGTGACTTCATCAAGCAACCCCATGACATG 1215
 QY 401 SerThrIleLysSerLysLeuGluAlaArgGluTyrArgAspAlaGlnGluPheGlyVala 420
 DB 1216 AGCAAAATCAAGTCTTAACTGAGAGCCGTGAGTACCTGATGCTCGAGAGTTGGTGTCT 1275
 QY 421 AspValArgLeuMetPheSerAsnGlyTyrLysTyrAsnProProAspHisGluValVala 440
 DB 1276 GACGTCGATGATGATGTTCTCCACTGTATATAGTACAACCTCTGTACATGAGCTGTG 1335
 QY 441 AlaMetAlaArgLysLeuGlnAspValPheGluMetArgPheAlaLysMetProAspGlu 460
 DB 1336 GCCATGGCGCCGCAAGCTCCAGAGATGTGTGAAATGGCTTTGCCAAGATGCCGAGAG 1395
 QY 461 ProGluGluProValValAlaValSerSerProAlaValProProProThrLysValVal 480
 DB 1396 CCGTAGGAGCCAGTGGTGGCTGTCTCCCGGCAAGTGCCCTCCACCAAGGTTGTG 1455

[illegible]

PA	(GENE-) GENE LOGIC INC.
XX	
PI	Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J,
DR	WPI; 2002-435328/46.
XX	
PT	Detecting granulocyte activation by detecting differential expression
PT	of genes associated with granulocyte activation, which serves as
PT	diagnostic markers that is useful for monitoring disease states and
PT	drug toxicity
XX	
PS	Claim 1; SEQ ID NO 231; 114pp; English.
XX	
CC	The invention relates to detecting (M1) granulocyte (GC) activation
CC	(GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC	DNA chip analysis as given in the specification, and comparing
CC	the expression level to an expression level in an unactivated
CC	GC, where differential expression of Gs is indicative of GCA.
CC	Also included are modulating (M2) GA by contacting GC with an agent
CC	that alters the expression of at least one gene in Gs; (2) screening (M3)
CC	for an agent capable of modulating GCA or an inflammation (especially
CC	chronic) in a tissue, an allergic response in a subject, exposure of a
CC	subject to a pathogen or sterile inflammatory disease using the
CC	gene expression profile; (3) detecting (M4) an inflammation (especially
CC	chronic) in a tissue, an allergic response in a subject, exposure of a
CC	subject to a pathogen or sterile inflammatory disease, by detecting the
CC	level of expression in a sample of the tissue of gene(s) from Gs, where
CC	the level of expression of the gene is indicative of inflammation;
CC	(4) treating (M5) an inflammation (especially chronic) or in a tissue,
CC	an allergic response in a subject, exposure of a subject to a pathogen
CC	or sterile inflammatory disease, by contacting a tissue having
CC	inflammation with an agent that modulates the expression of gene(s)
CC	from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
CC	modulating GA; M3 is useful for screening an agent capable of modulating
CC	GCA preferably in an inflammation in a tissue; M4 is useful for
CC	detecting an inflammation (especially chronic) in a tissue, an allergic
CC	response in a subject, exposure of a subject to a pathogen or sterile
CC	inflammatory disease (e.g. psoriasis, rheumatoid arthritis,
CC	glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
CC	reflexion injury, ARDS, adult respiratory distress syndrome,
CC	inflammatory bowel disease, Crohn's disease, ulcerative colitis,
CC	periodontal disease; also bacterial infection, viral infection,
CC	parasitic infection, protozoal infection, fungal infection and M5 is
CC	useful for treating one of the above conditions. The present
CC	sequence represents a gene differentially expressed in granulocytes.
CC	Note: The sequence data for this patent did not form part
CC	of the printed specification, but was obtained in electronic
CC	format directly from WIPo at
CC	ftp.wipo.int/pub/published_pct_sequences.
XX	
SS	Sequence 4664 BP; 1084 A; 1336 C; 1293 G; 951 T; 0 other;
XX	
SA	
XX	
SA	Alignment Scores:
XX	
XX	Pred. No.: 3,34e-82 Length: 4664
XX	Score: 1888.00 Matches: 404
XX	Percent Similarity: 67.08% Conservative: 81
XX	Best Local Similarity: 55.88% Mismatches: 142
XX	Query Match: 51.88% Indels: 97
XX	DB: 24 Gaps: 20
XX	
US-09-700-590A-22 (1-688) x ABR83660 (1-4664)	
OY	5 serGlyProGlyThrArgLeuArgAsnLeuProValMetGlyAspGlyLeuGlnThrSer 24
DB	1780 GCACGACACCGAGGAAAGATGCCAAACCCCTCTCTTGTATGAGGGCTTTGAGAGCCCC 1839
OY	25 GlnMetSerThrThrGlnAlaGlnAlaGlnProGlnProAlaAsnAlaAlaSerThrAsn 44
DB	1840 ACAATGGCTTCGGGCTGCTGCT--TTGCAACTTATACCCCTGCC-----AAC 1881
OY	45 ProProProProGlnThrSerAsnProAlaAsnLysProLysArgGlnThrAsnGlnLeuGln 64
DB	1882 CCACGACCCCCGGAGGCTGTCAATGCCAAAGAGCCGAGCGAGTTACCAACCGAGCTGGAA 1941

XX Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
 KM metastatic liver tumor; cytostatic; expression profile; disease state;
 KW disease progression; drug toxicity; drug efficacy; drug metabolism.
 OS Homo sapiens.
 PN WO200229103-A2.
 PD 11-APR-2002.
 PF 02-OCT-2001; 2001WO-US30589.
 PR 02-OCT-2000; 2000US-237054P.
 PA (GENE-) GENE LOGIC INC.
 PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
 DR WPI; 2002-426119/45.
 PT Diagnosing and detecting the progression of liver cancer,
 PT hepatocellular carcinoma or metastatic liver tumor in a patient,
 PT involves detecting the level of expression of two or more genes in a
 PT liver tissue sample -
 PS Claim 1: SEQ ID NO 1657; 298bp; English.
 XX The invention relates to a novel method for diagnosing and detecting the
 CC progression of liver cancer, hepatocellular carcinoma or metastatic liver
 CC tumor in a patient, and differentiating metastatic liver cancer from
 CC hepatocellular carcinoma in a patient, involving detecting the level of
 CC expression of two or more genes represented in ABN93503-ABN97455 in a
 CC tissue sample. The method of the invention has hepatotropic, and
 CC cytostatic activity. The method is useful for diagnosing and detecting
 CC the progression of liver cancer, hepatocellular carcinoma and metastatic
 CC liver carcinoma in a patient. The method is useful for identifying
 CC expression profiles which serve as useful diagnostic markers as well as
 CC markers that can be used to monitor disease states, disease progression,
 CC drug toxicity, drug efficacy and drug metabolism.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 SO Sequence 4664 BP; 1084 A; 1336 C; 1293 G; 951 T; 0 other;
 Alignment Scores:
 Pred. No.: 3,34e-82 Length: 4664
 Score: 1888.00 Matches: 404
 Percent Similarity: 67.08% Conservative: 81
 Best Local Similarity: 55.88% Mismatches: 142
 Query Match: 51.88% Indels: 97
 Gaps: 20
 US-09-700-590A-22 (1-688) x ABN95159 (1-4664)
 QY 5 SerGlyProGlyThrArgLeuArgAsnLeuProValMetGlyAspGlyLeuGluThrSer 24
 DB 1780 GCGACGACGAGGAAAGATTCGAAACCCCTCTTGTATGAGGGCTTTGAGAGCCCC 1839
 QY 25 GlMetSerThrThrGlnAlaGlnAlaGlnProGlnProAlaAsnAlaIAserThrAsn 44
 DB 1840 ACAATGGCTTGGTCTGCT---TTGCAACTTACCCCTGCC-----AAC 1881
 QY 45 ProProProGluThrThrSerAsnProAsnLysProLysArgGlnThrAsnGlnLeuGln 64
 DB 1882 CCAACACCCCGGAGGTGTCATCCCAAAAGCAGAGAGATTACCAACAGCTGCAA 1941
 QY 65 TyrLeuLeuArgValAlaLeuLysThrLeuTrpLysHisGlnPheAlaTrpProPheGln 84
 DB 1942 TACCTTACACAGAGTAGTGAAGGCTCTGTGAAACATCAGTTCCGATGCCATTCGCG 2001
 QY 85 GlnProValAspAlaValLysLeuAsnLeuProAspTyrTrpLysIleIleLysThrPro 104

DB 2002 CAGCCCTGTGAATGCTGTAACCTGGCTTACCGGATTATCACAAAATTTAAACAGCCT 2061
 QY 105 MetAspMetGlyThrIleLysLysArgLeuGluAsnAsnTyrTrpAsnAlaGlnGlu 124
 DB 2062 ATGACATGGGACTATTTAAGAGAGACTTGAACAAATATTATTATGGCTGCTCAGAG 2121
 QY 125 CysIleGlnAspPheAsnThrMetPheThrAsnCysTyrIleTyrAsnLysProGlyAsp 144
 DB 2122 TGTATGCAAGATTTTATATCCATGTTACCAACCTGTTACATTATACAAAGCCACTGAT 2181
 QY 145 AspIleValLeuMetAlaGluAlaLeuGluLysLeuPheLeuGlnLysIleAsnGluLeu 164
 DB 2182 GATATTGCTCTAATGACCAACACGCTGGAAGAAAGATTCTTACAGAGGTTGCATCAAG 2241
 QY 165 ProThrGluGluThrGluIleMetIleValGlnAlaLysGly-----ArgGly--- 180
 DB 2242 CCAAGAAAGAACAGAGCTGTAGTACCATTCCTTAAGAACACCCACAAAGAGGGGGCC 2301
 QY 181 -----ArgGlyArgLysGluThrGlyThrAlaLysProGlyLysThr 195
 DB 2302 AAGTTGAGCGCGCTCCAGGAGGTGTACAGTCCCATCAGGTGCTCGCTCTTCT 2361
 QY 196 ValProAsnThrThrGlnAlaSerThrProProGlnThrGlnThrProGlnProAsnPro 215
 DB 2362 GTGTACACACAGCCCTGTATCTCTCCACCTGAGATA-----CCT 2403
 QY 216 ProProValGlnAlaThrProHisProHisProAlaValThrProAspLeuIleValGln 235
 DB 2404 ACCAGTCTCTCAACATTCGCCACCA-----TCAGCATTTTCC 2442
 QY 236 ThrProValMetThrValAlaProProGlnProLeuGlnThrProProValProPro 255
 DB 2443 TCTCCACTTCTCAAGCTTGAC-----TCGTGAGACCCCGGCTCTGCT 2490
 QY 256 GlnProGlnProProAlaProAlaProGlnProValGlnSerHisProProIleIle 275
 DB 2491 GTTACTGACAGCTCTCCACGCC----- 2511
 QY 276 AlaAlaThrProGlnProValLysThrLysGlyValLysArgLysAlaAspThrThr 295
 DB 2512 -----CAGCCCTGTGCAAGAAAGAGGGTAAAGCGGAACAGATCTACCC 2559
 QY 296 ThrProThrThrIle-----AspProIleHisGluProProSerLeu 309
 DB 2560 ACCCTTACACCTACAGCCATCTTGCTCTGCTGCTTCCAGCTACGCCCTCTGGAGATCTT 2619
 QY 310 ProProGluProLysThrThrLysLeu---GlyGlnArgArgGluSerSerArgProVal 328
 DB 2620 -----GAGCTTAAGGACACAGCGCTTCCCTATGCGTAGAGAGGTGTCGCCCATC 2673
 QY 329 LysProProLysLysAspValProAspSerGlnGlnHisProAlaProGluLysSerSer 348
 DB 2674 AAGCCCCCAGCAGCAAGATCTGCTGCTCTCAGCAACACACAGAGCTCTAAGAAAGGA 2733
 QY 349 LysValSerGluGlnLeuLysCysCysSerGlyIleLeuLysGluMetPheAlaLysLys 368
 DB 2734 AAGCTTTCAGAACAGTTAAACCTTGCAATGCGATTTTGAAGAGTTACTCTTAAGAG 2793
 QY 369 HisAlaAlaTyrAlaTrpProPheTyrLysProValAspValGluAlaLeuGluLysHis 388
 DB 2794 CATGCGCCATATGCTTGCTTCTATAAACAGTGAAGTCTCTGACACTTGCGCAT 2853
 QY 389 AspTyrCysAspIleIleLysHisProMetAspMetSerThrIleLysSerLysLeuGlu 408
 DB 2854 GACTACCATGACATCTTAAGACCCCATGACCTTGAAGCTGTCAAGCGAGATGAG 2913
 QY 409 AlaArgGluTyrArgAspAlaGlnGluPheGlyAlaAspValArgPheMetPheSerAsn 428
 DB 2914 AACCGTGAATTAACGGGATGACAGGAGTTTCTGCTGATGATACGCTTATGTTCTCAAC 2973
 QY 429 CysTyrLysTyrAsnProProAspHisGluValAlaMetAlaArgLysLeuGlnAsp 448

Db 2974 TGCTAATAGTAAATCCCCAGATCAGATCTTGTGGCAATGGCAAGAAAGCTACAGAT 3033
 Qy ValPheGluMetArgPheAlaLysMetProAspGluProGluGluProValAlaVal 468
 Db 3034 GTATTGAGTTCCCTTATGCCAAGATGCAATGACATGACACCAAGACAGGCGCTTACCA 3093
 Qy SerSerProAlaValProProPro---ThrLysValValAlaProProSerSerAsp 487
 Db 3094 GTCTCTACTGCGCATGCCCTCGGCTTGGCCAAATCGTCTTCAGAGTCTCCAGTAGAGAA 3153
 Qy SerSerSerAspSerSerSer-----Asp 495
 Db 3154 AGTAGCGGTGAGAGCTCTCTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 3213
 Qy SerAspSerSerThrAspAspSerGluGluGluGluGluGluGluGluGluGluGluGlu 515
 Db 3214 GAAGAGAGTGAAGAGCTGAGATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3273
 Qy Glu 535
 Db 3274 GAACAGCTTCGGGAGTACATGAACAACAGCTGCTGTCTCCAGAGGTCCAAATATCCAAAG 3333
 Qy ProLysLysLysGluLysAspLysLysLysLysLysLysLysLysLysLysLysLys 551
 Db 3334 CCCAAGAGGAAA---AGAT 3390
 Qy LysArgLysGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 569
 Db 3391 CGAGGCGGAGTGGGGCGGATGAGATGACAAAGGGGCTTACGGGCGCCGCCACCTCAA 3450
 Qy ProLysLysThrLysLysAsnAsnSerSerSerSerSerSerSerSerSerSerSerSer 589
 Db 3451 CCTAAGAAAGTCCAGAAAGCAAGTGGCAGTGGGGGTGGCAGTGTCTTACGGCTTCT 3510
 Qy 590 -----ProMetLysSerLys 594
 Db 3511 GGGCTTTGAGCTTTCGAGAGAGTGGCACCAAGCTCCCAAAAAGGCCAACAAAGACGCC 3570
 Qy ProPro-----ProThr---TyrGluSerGluGluGluGluGluGluGluGluGluGluGlu 611
 Db 3571 CCACCTGCGCCCTCCTACAGCTTATCATTCACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3630
 Qy TyrGlu 631
 Db 3631 TACGATGAGAGCGGCGACGCTGAGCATCAACAATTACCTGGGGAGAGAGAGCTGGGC 3690
 Qy 632 ArgValValHisIleIleGlnSerArgLysProSerLysLysSerSerProAspGlu 651
 Db 3691 CGAGTTGTGCATATAATCAAGCGAGAGCGCTCTTACGATTCAAACCCAGAAAGAG 3750
 Qy IleGluIleAspPheGluThrLeuLysProSerThrLeuArgGluLeuGluValLeuGlu 671
 Db 3751 ATTGAGATTGATTGTAACACTCAAGCCATCCACACTTAGAGAGCTTGA-GCGGTATGT 3809
 Qy 672 HisLeuLeu 674
 Db 3810 CCTTCTCTG 3818
 RESULT 4
 AAV68343
 ID AAV68343 standard; cDNA to mRNA; 3104 BP.
 AC AAV68343;
 XX 05-MAR-1999 (first entry)
 DE Transcriptional regulatory factor RING3 encoding cDNA.
 XX
 KW Human; transcriptional regulatory factor; RING3; TSB; cancer;
 KW testis specific bromodomain; testicular cell proliferation; ds.
 XX
 OS Homo sapiens.
 XX

FH Key Location/Qualifiers
 FT CDS 106..2949
 FT /*tag= a
 PN MO9848015-A1.
 XX
 PD 29-OCT-1998.
 XX
 PF 17-APR-1998; 98MO-DP01782.
 XX
 PR 18-APR-1997; 97JP-0116402.
 XX
 PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
 XX
 PI Jones MH;
 XX
 DR WPI; 1998-583658/49.
 DR P-PSDB; AAW81168.
 XX
 PT Transcriptional regulator gene containing bromodomain sequence - may
 PT be expressed in testis tissue and is useful in treatment of cancer
 PT and other proliferative disorders
 XX
 PS Claim 2; Page 24-31; 42pp; Japanese.
 XX
 CC The present sequence encodes the human transcriptional regulatory
 CC factor RING3, which is isolated from testicular cells. RING3 contains
 CC a testis specific bromodomain (TSB) which is expressed specifically
 CC in testis tissue and also expressed in certain tumour lines. The
 CC transgenic cells may be used to express RING3 which is a TSB expression
 CC protein. The TSB expression product can be used in the treatment of
 CC cancer and other proliferative disorders, and in screening of compounds
 CC for ability to bind to it (e.g. for use as drugs by modulation of
 CC transcriptional regulation). DNA capable of hybridising to RING3
 CC polynucleotides may be used for construction of probes and primers.
 XX
 SQ Sequence 3104 BP; 1179 A; 560 C; 593 G; 772 T; 0 other;
 Alignment Scores:
 Pred. No.: 8.8e-64 Length: 3104
 Score: 1501.50 Matches: 326
 Percent Similarity: 61.67% Conservative: 81
 Best Local Similarity: 49.39% Mismatches: 160
 Query Match: 41.26% Indels: 94
 DB: Gaps: 13
 US-09-700-590A-22 (1-688) x AAV68343 (1-3104)
 Qy 35 ProGluProAlaAsnAlaAlaSerThrAsnProProProGluThrSerAsnProAsn 54
 Db 115 CCAAGTGCACAAACAGCTATTATTGTTTAACTTCTCCACCAAGATATATATTAATAAG 174
 Qy LysProLysArgGluThrAsnGluGluGluGluGluGluGluGluGluGluGluGluGlu 74
 Db 175 AAAAATGGCGCATTTGACAAATCAACTTCAGTATTCACAAAAGTTGCTTAAAGATTTA 234
 Qy 75 TrpLysHisGlnPheAlaTrpProPheGlnGluProValAspAlaValLysLeuAsnLeu 94
 Db 235 TGAAGCATTAAGTTTTCATGCGCCCTTCAACGTCCTGATGCTGGAACCTAAAGTTG 294
 Qy 95 ProAspTyrTyrLysIleIleLysThrProMetAspMetGluThrIleLysLysArgLeu 114
 Db 295 CCTGATTATTAATCAATTAATAAAACCACATGATTTAATAATTAAGAACCGCTTG 354
 Qy 115 GluAsnAsnTyrTyrTrpAsnAlaGlnGluCysIleGlnAspPheAsnThrMetPheThr 134
 Db 355 GAGATATAATATTATGCGAAGGCTTCAGAAATATAGAGCTTCATATACATATTTCTCA 414
 Qy 135 AsnCysTyrIleTyrAsnLysProGluAspAspIleValLeuMetAlaGluAlaLeuGlu 154
 Db 415 AATTGTTATTATATTAACCAAGCTCGAGATGACATTTGCTTAAGGCAACAAGCTTAAG 474
 Qy 155 LysLeuPheLeuGluLysIleAsnGluLeuProThrGluGluThrGluIleMetIleVal 174

D	475	AAGCTGTTATGAGAAATATATCTCAATGCCACAGAAAGCAACTTGGGTGTT---	531
Q	175	GlnAlaLysGlyArgGlyArgGlyGluThrGlyThrAlaLysProGluValSer	194
D	532	-----AAGAAAGAAATCAAGAAAGGCACT-----	555
Q	195	ThrValProAsnThrThrGlnAlaSerThrProProGlnThrGlnThrProGlnProAsn	214
D	556	-----CAACAGAAAT-----	564
Q	215	ProProProValGlnAlaThrProHisProPheProAlaValThrProAsnLeuIleVal	234
D	565	ATAGCTGTTCTTGTGTTAAAGAAAAATCTCAACCCAGCCCAACAGAAAAAGATTATTAAG	624
Q	235	GlnThrProValMetThrValValProProGlnProLeuGlnThrProProProValPro	254
D	625	CAGAGAAATCTCTCTGTAATTTCCTAAGACATCATTT-----	663
Q	255	ProGlnProGlnProProProProAlaProAlaProGlnProValGlnSerHisProProIle	274
D	664	-----TCTCCCTTGAACCTGGTATACAGGAGCACTTCAGTC	696
Q	275	IleAlaIaThrProGlnProValLysThrLysGlyValAlaLysArgValaAspThr	294
D	697	AACCTCAGTTCAAACTGCGGCCCAAGTTTACAAAGGTGTGAAGGAAACAGATTTACA	756
Q	295	ThrThrProThrThrIleAspProIleHisGluProProSerLeuProProGluProLys	314
D	757	ACAACCTCTGCACT---TCAGCAGTTAAAGCAAGATGAAATTTCTTCA-----	804
Q	315	ThrThrLysLeuGlyGlnArgGluSerSerArgProValLysPro-----ProLys	332
D	805	-----ACATTCACAGAAAAATACAGTGGCACTGCCCATTAATAAAGAAAAATATGCCAAAG	858
Q	333	LysAspValProAspSerGlnGlnHisProAlaProGluLysSerSerLysValSerGlu	352
D	859	AATGTTTGGCCAGATTCTCAGCAACAATATATATGTTGTGAGCTGTTAAAGTAACTGAA	918
Q	353	GlnLeuLysCysCysSerGlyIleLeuLysGluMetPheAlaLysLysHisAlaIaThr	372
D	919	CAATTAAAGCAGCTGATGAGATTTCTTAAAGAAATCTTGCAAGAAATATTTTTCATAT	978
Q	373	AlaTrpProPheTrpLysProValAspValGluAlaLeuGlyLeuHisAspTrpCysAsp	392
D	979	GCATGGCCCTTTTATATCCGTTGATGAGTTAAAGCTTTGGAACTCCATATCACTATATGAC	1038
Q	393	IleIleLysHisProMetAspMetSerThrIleLysSerLysLeuGluAlaArgGluTrp	412
D	1039	GTTTGTCAAAAATCCGATGGATCTTGGAACTATTAAAGAGAAATGATTAACCAAAATAT	1098
Q	413	ArgAspAlaGlnGluPheGlyAlaAspValArgLeuMetPheSerAsnTrpTrpLysTrp	432
D	1099	AAGAGTGCATTCATTTGGCGGAGATGTAGTTATGTCTATGTAATGTCTCAAAAGTAC	1156
Q	433	AsnProProAspHisGluValAlaAlaMetAlaArgLysLeuGlnAspValPheGluMet	452
D	1159	AATCTCTCAAGTACGAAAGTTGTGACATGCGAAGATGCTTCAGATGTTTTTCCAAAGC	1218
Q	453	ArgPheAlaLysMetProAspGluProGluGln-----ProValValAlaValSerSer	470
D	1219	CATTTCCTCAAGAGCCGATGTAACCTGTTGAAGATGCTTTATGTTACATCAACAAACA	1278
Q	471	ProAlaValProProProThrTrpValValAlaProProSerSerSerAspSerSerSer	490
D	1279	GAT-----ATACAGAAACCACTGGTATGAGAAACACTAAT	1314
Q	491	AspSerSerSerAspSerAspSerSerThrAspAspSerGluGluGluArgIaGlnArg	510
D	1315	GAAAGCTCTCTGAAGGAACTCTTC---GATGATTTCTGAAGATGACGGAATTAAAGGT	1371
Q	511	LeuAlaGluLeuGlnGluGlnLeuLysAlaValHisGluGlnLeuAlaIaLeuSerGln	530

Db	13732	CTTCGAAGGCTTAGAGGACGCTTAAAGCTGTACATCAACAGCTCCAGGTTTGTCCAA	1431
Oy	531	---ProGlnGlnAsnLysProLysLysLysGlnLysAspLysLysGlnLysLysGln	549
Db	1432	GTACCTTTCCGTAAAGCTTAATAAAAAGAAAGAGAGCTTAAAAAGGAAAAAGAA	1491
Oy	550	Lys-----HisLysArgLysGlnLysLysLysGlnLysLysSer	563
Db	1492	AAGGTTAATTAACAGCAATGAAAAATCCAGAAAAATGTGTGAGCAATGAGGCTTAAAGAA	1551
Oy	564	LysLysLysGlnProProProLysLysThrLysLysAsnAsnSerSerAsnVal	583
Db	1552	AAGTCCAAAGGAAGAAATCAGCCAAAGAAAAAGAAACAACGCTTCATTGGT-----	1599
Oy	584	SerLysLysGlnProLysLysLysSerLysProProProThrThrLysGlnSerGln	603
Db	1600	-----CTAAATCTGAAGAT	1614
Oy	604	GlnAspLysCysLysProMetSerLysGlnLysLysLysLysLysLysLysLysLys	623
Db	1615	GAAATATATCTTAAACCTATGAACTATGATGAGAAAGGAGTTAACTGAATATTAAC	1674
Oy	624	LysLeuProGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys	643
Db	1675	AAATCTCTGAGATTAACCTTGCGGAGAGTTCACATTAATACATCAAGAGACCTTCT	1734
Oy	644	LeuLysAsnSerAsnProAspGlnLysLysLysLysLysLysLysLysLysLysLys	663
Db	1735	CTGAGCAATTCGAATCTGATGAGATAGAGATAGACTTTGAAACACTGAAAGCATCAAC	1794
Oy	664	LeuArgGlnLeuGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLys	683
Db	1795	CTAAGAGAAATTAAGA-AAAATATATGTTTCGCGCAGT-----TCTAAGAAAAAGACCATTTAA	1847
RESULT 5			
ID	AAK40172	standard; DNA; 3106 BP.	
XX	AAK40172;		
AC	AAK40172;		
DT	02-JUL-1999	(first entry)	
XX			
DE	W09904265	Seq ID No: 686.	
XX			
KW	cancer associated antigen; diagnosis; research; treatment; human; breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer; prostate cancer; ss.		
KW			
KW			
KW			
OS	Homo sapiens.		
XX			
PN	W09904265-A2.		
XX			
BD	28-JAN-1999.		
PF	15-JUL-1998;	98WO-US14679.	
XX			
XX	22-JUN-1998;	98US-0102322.	
PR	17-JUL-1997;	97US-0896164.	
PR	10-OCT-1997;	97US-0061599.	
PR	10-OCT-1997;	97US-0061765.	
PR	10-OCT-1997;	97US-0948705.	
PR	11-OCT-1997;	97GB-0021697.	
XX			
PA	(LUDW-) LUDWIG INST CANCER RES.		
XX			
PI	Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;		
PI	Pfeundtschuh M, Sahin U, Scanlan MJ, Stockert E;		
PI	Tureci O;		
XX			
DR	WPI; 1999-132448/11.		
XX			

PT for the diagnosis, monitoring or treatment of cancers
 XX
 XX Claim 67, Page 730-731, 787pp; English.

CC The invention relates to a method for diagnosing a disorder characterised
 CC by expression of a human cancer associated antigen precursor coded for by
 CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
 CC biological sample isolated from a subject with an agent that specifically
 CC binds to the NAM, an expression product or a fragment of an expression
 CC product complexed with an HLA molecule; and (b) determining the
 CC interaction between the agent and the NAM or the expression product as a
 CC determination of the disorder. The products and methods can be used in
 CC the diagnosis, monitoring, research, or treatment of conditions
 CC characterised by the expression of various cancer associated antigens.
 CC The invention provides nucleic acid sequences and encoded polypeptides
 CC which are cancer associated antigen precursors expressed in human breast
 CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
 CC lung cancer.

CC Sequence 3106 BP; 1179 A; 560 C; 594 G; 773 T; 0 other;

Alignment Scores:

Prod. No.: 8-8e-64 Length: 3106
 Score: 1501.50 Matches: 326
 Percent Similarity: 61.67% Conservative: 81
 Best Local Similarity: 49.39% Mismatches: 160
 Query Match: 20 Gaps: 13

US-09-700-590A-22 (1-688) x AAY40172 (1-3106)

QY 35 ProGlnProAlaAsnAlaAlaSerThrAsnProProGluThrSerAsnProAsn 54
 Db 117 CCAAGTGCACAACAGCTATATTGTTAACTCTCCACCGAATATATTAATCTAAG 176
 QY 55 LysPProLysArgGlnThrAsnGlnLeuGlnIleuLeuValValLeuLysThrLeu 74
 Db 177 AAAATATGGCCATTCGACCAATCACTCATCAAAAAGTTGCTCTAAGCATTTA 236
 QY 75 ThrLysHisGlnPheAlaThrProPheGlnGlnProValAlaAlaValLysLeuAsnLeu 94
 Db 237 TGGAAAGCTATGTTTCTTTCATGGCCCTTCAACGCTCTGGAGATGCTGTAAACTTAAGTTG 296
 QY 95 ProAspTyrTyrLysIleIleLysThrProMetAspMetGlyThrIleLysLysArgLeu 114
 Db 297 CCTGATTATTATACCATTTATAAAAAACCAGTGCATTTAAATACATTAAGAAAGCGCTTG 356
 QY 115 GluAsnAsnTyrTyrTrpAsnAlaGlnGlnIleGlnIleGlnAspPheAsnThrMetPheThr 134
 Db 357 GAGATAATATATTATGCGAAGGCTTCAGAAATGTATAGAAAGCTTCATATCAATGTTCTCA 416
 QY 135 AsnGlyTyrIleTyrAsnLysProGlyAspAspIleValIleMetAlaGlnAlaLeuGln 154
 Db 417 AATTGTTATTTATTAACAGCTGAGATGACATGTTCTTATGGCACAGCTCTTAAG 476
 QY 155 LysLeuPheLeuGlnLysIleAsnGlnLeuProThrGlnGlnIleMetIleVal 174
 Db 477 AAGCTGTTTATGCAAAATTTATCTCAGATGCCACAAGAAAGCAAGTTGTGGCTTT--- 533
 QY 175 GlnAlaLysGlyArgGlyArgLysGlnThrGlyThrAlaLysProGlyValIser 194
 Db 534 -----AAGGAAGAATTCAGAAAGGACT----- 557
 QY 195 ThrValProAsnThrThrGlnAlaSerThrProProGlnThrGlnThrProGlnProAsn 214
 Db 558 -----CAACAGAAAT 566
 QY 215 ProProProValGlnAlaThrProHisProPheProAlaValThrProAspLeuIleVal 234
 Db 567 ATAGCTGTTCTTCTGCTTAAGAAAAATCATCATCCAGCGCAACGAAAAAGTATTTAAG 626
 QY 235 GlnThrProValMetThrValValProProGlnProLeuGlnThrProProProValPro 254
 Db -----CTAAATCTGAAGAT 1616

Db 627 CAGCAAGAAATTCCTTGTGATTTCCTAAGACATCTATT----- 665
 QY 255 ProGlnProGlnProProProAlaProAlaProGlnProValGlnSerHisProProIle 274
 Db 666 -----TCTCCCTTGAACGTGGTACAGGGAGCTTACGTC 698
 QY 275 IleAlaAlaThrProGlnProValLysThrLysLysGlyValLysValGlyValAlaSerThr 294
 Db 699 AACCCAGCTTCACAAACTGCGGCCCAAGTTACAAAGAGTGTGAAGAGAAAGCAGATACA 758
 QY 295 ThrThrProThrThrIleAspProIleHisGlnProProSerLeuProProGluProLys 314
 Db 759 ACAACTCTCTGCACT---TAGCAGTTTAAAGCAAGTAACTTTTCTCCA----- 806
 QY 315 ThrThrLysLeuGlnGlnArgArgGlnLysSerSerArgProValLysPro-----ProLys 332
 Db 807 -----ACATTACAGAAAAATCAGTGGACCTGCCACTCTTAAAGAAATATGCCAAG 860
 QY 333 LysAspValProAspSerGlnGlnHisProAlaProGluLysSerSerLysValSerGlu 352
 Db 861 AATGTTTGGCCAGATTCCTCAGCAACAATATATGTTGTGAGACTGTTAAGTAAGTAACTGAA 920
 QY 353 GlnLeuLysCysCysSerGlyIleLeuLysGlnMetPheAlaLysLysHisAlaAlaTyr 372
 Db 921 CAATTAAAGCAGCTGTATGAGATCTTAAAGAAATGCTTGCAAAAGAAACATTTTCTAT 980
 QY 373 AlaTrpProPheTyrLysProValAspValGlnAlaLeuGlnLysHisAspTyrCysAsp 392
 Db 981 GCATGGCCCTTTTATATCTGTTGACGTTAATGCTTTGGGACTCCCATACTATATGAC 1040
 QY 393 IleIleLysHisProMetAspMetSerThrIleLysSerLysLeuGlnAlaArgGluTyr 412
 Db 1041 GTGTCAAAAATCCGATGAGATCTTGGAATTAAGAGAAATAGTAACTCAAGAAATAT 1100
 QY 413 ArgAspAlaGlnGlnPheGlyAlaAspValArgLeuMetPheSerAsnGlyTyrLysTyr 432
 Db 1101 AAGAGTGATATCTATTTCGGCGGAGATGTTAGATTAAATGTTCAATATGCTCAAGATAC 1160
 QY 433 AsnProProAspHisGlnValAlaValAlaMetAlaArgLysLeuGlnAspValPheGluMet 452
 Db 1161 AATCCTCCAGATCACGAAGTTGTGACAAATGCGCAAGATGCTTCAGATGTTTGGAAAGC 1220
 QY 453 ArgPheAlaLysMetProAspGlnProGlnGln-----ProValValAlaValSerSer 470
 Db 1221 CATTTTCAAGAGATCCGATGGAACCTGTGAGAGTATGCTTATATGTCATCAATAACA 1280
 QY 471 ProAlaValProProProThrLysValValAlaProProSerSerSerSerSerSerSer 490
 Db 1281 GAT-----ATCACAGAAACCACTGGTAGAGAGACATTAAT 1316
 QY 491 AspSerSerSerAspSerAspSerSerThrAspAspSerGlnGlnIleValArgAlaGlnArg 510
 Db 1317 GAAGCTCTCTGAAAGGAAGACTTCTT---GATGATTTCTGAAGATGACCGAGTTAAAGCGT 1373
 QY 511 LeuAlaGlnLeuGlnGlnGlnLeuLysAlaValHisGlnGlnLeuAlaAlaLeuSerGln 530
 Db 1374 CTTGCAAAAGCTTCAGAGAGCTTAAAGCTGTACATCAACAGCTCCAGGTTTGTCCCAA 1433
 QY 531 ---ProGlnGlnAsnLysProLysLysLysGlnLysAspLysLysGlnLysLysLysGln 549
 Db 1434 GTACCTTTCCTGTAAGCTTAATTAAGAAAGAGAGCTTAAGAAAGAAAGAAAGAAAGAA 1493
 QY 550 Lys-----HisLysArgLysGlnGlnValGlnGlnLysLysSer 563
 Db 1494 AAGGTTAATACAGCAATGAATAATCCAAAGAAATGTGTGAGCAAAAGAGCTTAAAGAA 1553
 QY 564 LysAlaLysGlnProProProLysLysThrLysLysAsnAsnSerSerAsnSerAsnVal 583
 Db 1554 AAGTCCAAAGAAATCAAGCCAAAGAAAGAAACAACAGTTCAATGCT----- 1601
 QY 584 SerLysLysGlnProAlaProMetLysSerLysProProProThrTyrGlnSerGlnGln 603
 Db 1602 -----CTAAATCTGAAGAT 1616

QY 604 GluAspLysCysLysProMetSerTyrGluGluLysArgGlnLeuSerLeuAspIleAsn 623
 DB 1617 GAAGATAATCTCTAACCTATGAGATGAGAGAAAAGCAGTTAAGTCGAAATATTAAC 1676
 QY 624 LysLeuProGluLysLeuGluYargValValHisIleIleGlnSerArgGluProSer 643
 DB 1677 AAACCTCCCTGGAGATAAATCTTGGCGAGTTCACATAATACATCAAGAGCCCTTCT 1736
 QY 644 LeuLysAsnSerAsnProAspGluIleGluIleAspPheGluThrLeuLysProSerThr 663
 DB 1737 CTAGACAAATTCATCTCTGATGAGATGAGATGACCTTTAAACACCTGAAAGCATCAACA 1796
 QY 664 LeuArgGluLeuGluYalaLeuCysHisLeuLeuPheAlaGluGluYargLysGluThrPheLys 683
 DB 1797 CTAGAGAAATTAGA-AAATATCTTTCGCGCATG-----TCTAAGAAAGAGACCATTA 1849

RESULT 6
 AAH18591
 ID AAH18591 standard; cDNA; 1884 BP.
 AC AAH18591;
 XX
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human cDNA sequence SEQ ID NO:18782.
 XX
 XX Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.
 OS Homo sapiens.
 XX
 PN EP1074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 PF 28-JUL-2000; 2000EP-0116126.
 XX
 PR 29-JUL-1999; 99JP-0248036.
 XX
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 XX (HELI-) HELIX RES INST.
 PA
 PI Ota T, Isegai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 DR WPI; 2001-318749/34.
 XX
 XX Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX
 PS Claim 8; SEQ ID 18782; 2537bp + CD ROM; English.
 XX
 XX The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by

CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialized methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.

XX Sequence 1884 BP; 492 A; 536 C; 451 G; 405 T; 0 other;
 SQ

Alignment Scores:
 Pred. No.: 1,656-61 Length: 1884
 Score: 1450.00 Matches: 308
 Percent Similarity: 65.78% Conservative: 61
 Best Local Similarity: 54.90% Mismatches: 115
 Query Match: 39.85% Indels: 77
 DB: 22 Gaps: 14

US-09-700-590A-22 (1-688) x AAH18591 (1-1884)

QY 16 ValMetGlyAspGlyLeuGluThrSerGlnMetSerThrThrGlnAlaGlnAla----- 33
 DB 330 ATCAGTCGCGCGCATCTGGACACCCCTGCGTGTCTCAGAGACACATCCCTCGGG 389
 QY 34 ---GlnProGlnProAlaAsnAlaAlaSerThrAsnProProProGluThrSerAsn 52
 DB 390 GGCCAATGTGTGCTGTGTGTGCGATTCAAAC-----CCGAGGTGTCCAAT 437
 QY 53 ProAsnLysProLysArgGlnThrAsnGlnLeuGlnTyrLeuLeuArgValValLeuLys 72
 DB 438 CCCAAGAGCCGAGAGAGAGTTCACCAAGCTGCAATACCTACACAGAGTAGATGAAG 497
 QY 73 ThrLeuTrpLysHisGlnPheAlaTrpProPheGlnGlnProValAspAlaValLysLeu 92
 DB 498 GCTGTGTGAAACATCATGTCATCCATGCCATTCGCGACCTGTGATGCTGTCAAACTG 557
 QY 93 AsnLeuProAspTyrTyrLysIleIleLysThrProMetAspMetGlyThrIleLysLys 112
 DB 558 GGTCTACCGGATTCACCAAAATTAATAAAGCCATATGAGATGAGTCTATTAAAGAG 617
 QY 113 ArgLeuGluAsnAsnTyrTyrTrpAsnAlaGlnGluCysAlaGlnAspPheAsnThrMet 132
 DB 618 AGACTTGAAGAACATATTATTATGGCTGCTTCAGAGTGTATGACAGATTATTATACCATG 677
 QY 133 PheThrAsnCysTyrIleTyrAsnLysProGlyAspAspIleValLeuMetAlaGlnAla 152
 DB 678 TTCACCACTGTTCATTTACACCAAGCCACTGATGATGATGCTTATGACACAAAG 737
 QY 153 LeuGluLysLeuPheLeuGlnLysIleAsnGluLeuProThrGluGluThrGluIleMet 172
 DB 738 CCGAAGAAAGATATTCTTACAGAGGTTGCATCATGCCACAGAAAGAACAGAGCTGTA 797
 QY 173 IleValGlnAlaLysGly-----ArgGly-----ArgGlyArg 183
 DB 798 GTGACCATCTCTAAGAACAGCAAGAGGGGGCCCAAGTTGGACGCTCAGGGCACT 857
 QY 184 LysGluThrGlyThrAlaLysProGlyValSerThrValProAsnThrThrGlnAlaSer 203
 DB 858 GTTACACAGGCCCATCAGAGTGCCTGCTCTCTCTGTCTACACACACACCCCTGATACT 917
 QY 204 ThrProGlnThrGlnThrProGlnProAsnProProProValGlnAlaThrProHis 223
 DB 918 CCTCCACCTGAGATA-----CTACACACGTCTCTCAACATTCCTCCAC 959
 QY 224 ProPheProAlaValThrProAspLeuIleValGlnThrProValMetThrValValPro 243
 DB 960 CCA-----TCAGCATTCCTCTCCACTTC----- 986
 QY 244 ProGlnProLeuGlnThr-----ProProProValProProGlnProGlnProPro-Pr 261
 DB 987 ---AGCTCTTGCACTCTGCTGAGACCCCGCTCTTGCTGTACACGACCTCTCCAGCC 1043
 QY 261 oAlaProAlaProGlnProValGlnSerHisProProIleIleAlaAlaThrProGlnPr 281
 DB 1043


```

Qy 128 AspPheAsnThrMetPheThrAsnCysTyrIleTyrAsnLysProGluYAspAspIleVal 147
Db 1272 GACCTCAATCGATGTTCAACAAATTGCTATGCTACAAACAGCCGCGAGATGATAGT 1331
Qy 148 LeuMetAlaGluIleValGluLysLeuPheLeuGlnLysIleAsnGluLeuProThrGlu 167
Db 1332 GTTATGCCCCAGACGCTCGGAAGGCTCTTCAGAAAGATCGAATGATGCTCTTAAGAG 1391
Qy 168 GluThrGluIleMetIleValAlaGlnAlaLysGluYArgGluYArgGluYArgGluYArg 187
Db 1392 GACCTGAACTGAGAGCCGCTGACAGGAGTGGCAAAAAGAAAGCAACGCGCGCGCT 1451
Qy 188 ThrAlaLys-----ProGluYAsnSerThrValProAsnThrThrGlnAla 202
Db 1452 ACCCCCAAGCATGTCGAGGTGGCGCTGAGACATCCACCGCTTCTGCTACATCCTCGCGCA 1511
Qy 203 SerThrProProGlnThrGlnThrProGlnProAsn-ProProProValGlnAlaThrPr 222
Db 1512 GCGATT-----ACGACGGAGCAGGTAGTGGCTCCACAAAGTGTGATGCTGCGC 1562
Qy 222 ohSProPhe-----ProAlaVal-----ThrProAspLeuIleVal 234
Db 1563 TCATCCGACACAAAGTCCGCTGCAAGAGCAACGAGGACAGCGGTGATCTTCAAGT 1622
Qy 234 GlnThrProValMetThrValValProProGlnProLeuGlnThrProProValPr 254
Db 1623 ACACCGGAACTCAGCGCGGTTCCGGTGGGAGAGAGCATGCGCGCCGACCGCATTC 1682
Qy 254 oPro-----GlnProGlnProProProAlaProAlaProGlnProValGln 269
Db 1683 GCCATGGCGCGCAGGTTTCATGACGCGCGCGGT-GCACGCTCCATACACCGATTAG 1741
Qy 269 nSerHisProPro-----GlnProGlnProProProAlaProAlaProGlnProValGln 273
Db 1742 CACAAATGCCACCGCACACGATACCCGAGCACCAATAGACAGACAGCATGCTGG 1801
Qy 273 ----- 273
Db 1802 CGCGTTGGTGGCGCAGGTGACAGTGGAGCCAAATCCAAATGTGCGGCTGTATGGCTAG 1861
Qy 273 ----- 273
Db 1862 TCTTCTGAATGCGGCGCAACGCGCGCTATCCCGGTGCGCCGCGCAGACGCGGTCAA 1921
Qy 273 ----- 273
Db 1922 TAGCTCTGCTTCTAGATGGCAGTACAGCCGAGTAGAGCGCGCAGACAGCGCTGC 1981
Qy 273 ----- 273
Db 1982 GCGGCGGCGCGCGCGCAGAGAGTGCAGCGGAGCGGCTGAGAGAGCAGAGCAATACC 2041
Qy 273 ----- 273
Db 2042 AGCAGTTGCGGTCAATGCGCGTAACGCGCTTACAGGCTATGATATCGGCGGTGAGCGT 2101
Qy 274 -----IleIleAlaIleThrPro-----GlnProValLysThrLysLysGluValLysAr 290
Db 2102 CGGAGTGAAGCGCGGTATACCGCTCAGCAGCGCGCCAAATATAAGAGGCTGCAACG 2161
Qy 290 GlyAlaAspThrThrThrProThrThrIleAspProIleIleGluProSerSerLeuPr 310
Db 2162 GAAAGGAGACAAACAAACGCGCAGCGC-----AATGCTTTGAATCCCGTACACGCA 2215
Qy 310 oProGluProLysThrThrLysLeuGluGlnArgGluSerSerArgProValLysPr 330
Db 2216 AATGACTCCAAATCGCCAAAGATTGCGACGCGCGGAGATCGAATCGT----- 2264
Qy 330 oProLysLysAspVal-----ProAspSerGlnGlnHisProAlaPro----- 344
Db 2265 -----CAGGATCTTACATTCCAGGGGCTCGGAGTACAAATATGTCGCGCTAGGCGTCTC 2317

```

```

Qy 345 -----GI 345
Db 2318 CGAGTGGCCCGGACATTGCGCGTCTACTGCGCGCGCGCTGGTGTGCTGGCGCA 2377
Qy 345 uLysSerSerLysValSerGluGlnLeuLysCysCysSerGluYleLeuYsgLumEPr 365
Db 2378 GAACAAGAGAAAGCTGCGATGCTGCTCAAGTGTCTCAACGAATCTTCAAGAGCTCTT 2437
Qy 365 eAlaLysLysHisAlaAlaTyrAlaThrProPheTyrLysProValAspValGluAla 385
Db 2438 CTCGAAGAAGCACTCGGCGCTATGCTGGCAATTCACAAAGCAGTGCAGCGGAAATGCT 2497
Qy 385 uGluLeuHisAspTyrCysAspIleIleLysHisProMetAspMetSerThrIleLysE 405
Db 2498 TGGCTCGCATGACTACACAGCATCATCAAGAAACCAATGATGTGGCAGCTCAAGCG 2557
Qy 405 LysLeuGluAlaArgGluTyrArgAspAlaGlnGluPheGluYAlaAspValArgLeuE 425
Db 2558 GAATATGACATCGGAGTACAGAGCGCGCGGAATTTCCCGCGAGTGGATTAAT 2617
Qy 425 cPheSerAsnCysTyrLysTyrAsnProProAspHisGluValAlaAlaMetAlaArgL 445
Db 2618 ATTCACCACTGCTACAAATGATACATCGGCAATCATGATGTTGTGGCATGGGTGCA 2677
Qy 445 sLeuGlnAspValPheGluMetArgPheAlaLysMetProAspGluProGluLnuProVa 465
Db 2678 GCTGCAAGAGCTATTGAGATGGCTATGCCAAATCCCGATGAGCCGATGCAATGC 2737
Qy 465 ValAlaValSerSerProAla----- 472
Db 2738 GGGCCACATACAGGACATGCGCGGACAGTGTATGGCCATGGCCATGCGTACCGGCA 2797
Qy 473 -----ValProProThrLysValAlaProProSerSerSerAspSers 489
Db 2798 TGCTACAGGACATGCTACATGCTACGCGATACCGCGCTCTCTCTCACTCAAGAC-GATG 2856
Qy 489 eSerAspSerSerSerAspSerAspSerSerThrAsp-----AspSerGluGlu 506
Db 2857 CCAGCGATTGCTCCAGCGAGACTCCAGCGATACCAAAACGAATCCATCGAGTAGAG 2916
Qy 506 LuArgAlaGlnArgLeuAlaGluLeuGlnGlnLeuLysAlaValHisGluGlnLeuA 526
Db 2917 AGCGAGCGCTAGGCTGAAGATGCTGAGTCAAGTCTGCTGCTGCGCAGAGGAATCC 2976
Qy 526 IaAlaLeuSerGlnProGlnGlnAsnLysProLysLysLysGluYAspLysGluL 546
Db 2977 GTAAGCTGTCAAGAGAGGCTCCGCCAAAGAGAGCAAGAAAACTCAGAGAAAAA 3036
Qy 546 YsLysLys----- 548
Db 3037 AAAAGTCATTAAGCGCGGATCAGGCTGTGCTGCGCTTCGACCATGTACAGCCACG 3096
Qy 548 ----- 548
Db 3097 GCGGCGGTCAATGCTGCGGAGAGAGCGGTCCCGATCAGCGGCGCATGGAGCGTTT 3156
Qy 548 ----- 548
Db 3157 CGGTGCGAGCGGTGCGGCTCTTGGGTCCCGGTGAGAGCGGCGCGCTAATCTCAAG 3216
Qy 548 ----- 548
Db 3217 CGGTGCGGTGCTCATTTGTTGGCCATGGGAGAGCGGCGGTGCGAGAGGGGTTCCA 3276
Qy 549 -----GluYsHisLysArgLysGluGluValGlu----- 558
Db 3277 AATGAGTGTCTTGACACACCGATTCAAGAGTGCACATGCGCTTTACAGATGGCGG 3336
Qy 558 ----- 558
Db 3337 GCGGCGGTGCGCGCGGCTGCTGCTTGTGTCGCGGTGACAGAGAGCATCGT 3396
Qy 558 ----- 558

```

Db 3397 CGGCGCGAAGCGGGGACACTGGCCGCTCTGCGAGCGGGCGAGCGGCGAGTGGCG 3456
 QY 559 -----GluAsnLysSerLysAlaLysGluProProProl 571
 Db 3457 GCGGTACAGCGGTGGCGAGCGGACAGTAAAGTGTCTAAGAGCAAGGCGGAGTGGCG 3516
 QY 571 YslYsThrLys-----LysAsnAsnSerSerAsnSerAsnVal 583
 Db 3517 CAAAGGCGAGCGGCGGCGGCGTGGTGGCGAATATATATATATATATATATATATATAT 3576
 QY 583 ----- 583
 Db 3577 CTGGTGGCGAGCGGAGCTGCGAGCGGCGCTGTGATCTGTGTGAGTGTGGCGGTGCGAG 3636
 QY 583 ----- 583
 Db 3637 GAGCAGCGGCGGCGGCGGCAATGCTTCAGAGCGGCGCAAGGCGAGCTTGGCGGCGCTG 3696
 QY 584 -----SerL 585
 Db 3697 GCGGTGGTGTGGTGGCGGCAATGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3756
 QY 585 ----- 585
 Db 3757 AGAAGAGACCTAGCGAGGTGATGAAC-----TTTGACTCCGAGGAGGAGG 3801
 QY 585 YslYsGluProAlaProMetLysSerLysProProProThrThrLysGluSerGluGluGlu 605
 Db 3802 ACACGCGCCAGCCATGTGTCATGATGAGAGCGGCGGCTGTGCTGCTGCTGCTGCTGCTG 3861
 QY 605 sPlYsCyLysProMetSerLysGluGluLysArgLysLysSerLysLysLysLysLysLys 625
 Db 625 euProGlyGluLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 645
 QY 645 YsaAsnSerAsnProAspGluLysLysLysLysLysLysLysLysLysLysLysLysLys 665
 Db 3922 GTGACTCCAAATCCCGATGAAATAGATGCACTTCGAGACCGCTGAGCGCTGCG 3981
 QY 665 rGgLUeugLYAlaLeuCYshLysLysLysLysLysLysLysLysLysLysLysLysLys 678
 Db 3982 GCGAGCTAGA-AAGCTATGTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4021
 RESULT 8
 ABK84162
 ID ABK84162 standard; cDNA; 1741 BP.
 AC ABK84162;
 XX
 DT 14-AUG-2002 (first entry)
 XX
 DE Human cDNA differentially expressed in granulocytic cells #733.
 XX
 KW Human; ss; granulocytic cell; DNA chip; bacterial infection;
 KW viral infection; parasitic infection; protozoal infection;
 KW fungal infection; sterile inflammatory disease; psoriasis;
 KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
 KW cardiac reperfusion injury; renal reperfusion injury; AIDS;
 KW adult respiratory distress syndrome; inflammatory bowel disease;
 KW Crohn's disease; ulcerative colitis; periodontal disease;
 KW granulocyte activation; chronic inflammation; allergy.
 KM
 XX
 OS Homo sapiens.
 XX
 PN W0200228999-A2.
 PD 11-APR-2002.
 XX
 PF 03-OCT-2001; 2001MO-US30821.
 PR 03-OCT-2000; 2000US-237189P.
 XX
 PA (GENE-) GENE LOGIC INC.

XX
 PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J,
 DR WPI; 2002-43528/46.
 XX
 PT Detecting granulocyte activation by detecting differential expression
 PT of genes associated with granulocyte activation, which serves as
 PT diagnostic markers that is useful for monitoring disease states and
 PT drug toxicity -
 XX
 PS Claim 1; SEQ ID No 733; 114pp; English.
 XX
 CC The invention relates to detecting (M1) granulocyte (GC) activation
 CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
 CC DNA chip analysis as given in the specification, and comparing
 CC the expression level to an expression level in an unactivated
 CC GC, where differential expression of Gs is indicative of GCA.
 CC Also included are modulating (M2) GA by contacting GC with an agent
 CC that alters the expression of at least one gene in Gs; (2) screening (M3)
 CC for an agent capable of modulating GCA or an inflammation (especially
 CC chronic) in a tissue, an allergic response in a subject, exposure of a
 CC subject to a pathogen or sterile inflammatory disease using the
 CC gene expression profile; (3) detecting (M4) an inflammation (especially
 CC chronic) in a tissue, an allergic response in a subject, exposure of a
 CC subject to a pathogen or sterile inflammatory disease, by detecting the
 CC level of expression in a sample of the tissue of gene(s) from Gs, where
 CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,
 CC an allergic response in a subject, exposure of a subject to a pathogen
 CC or sterile inflammatory disease, by contacting a tissue having
 CC inflammation with an agent that modulates the expression of gene(s)
 CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
 CC modulating GA; M3 is useful for screening an agent capable of modulating
 CC GCA preferably in an inflammation in a tissue; M4 is useful for
 CC detecting an inflammation (especially chronic) in a tissue, an allergic
 CC response in a subject, exposure of a subject to a pathogen or sterile
 CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis,
 CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
 CC reperfusion injury, AIDS, adult respiratory distress syndrome,
 CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,
 CC periodontal disease; also bacterial infection, viral infection,
 CC parasitic infection, protozoal infection, fungal infection and M5 is
 CC useful for treating one of the above conditions. The present
 CC sequence represents a gene differentially expressed in granulocytes.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 1741 BP; 481 A; 482 C; 408 G; 370 T; 0 other;
 XX
 Alignment Scores:
 Pred. No.: 2,54e-51 Length: 1741
 Score: 1237.50 Matches: 301
 Best Similarity: 60.33% Conservative: 61
 Percent Local Similarity: 50.17% Mismatches: 135
 Query Match: 34.01% Indels: 110
 DB: 24 Gaps: 15
 US-09-700-590a-22 (1-688) x ABK84162 (1-1741)
 QY 5 serGlyProGlyThrArgLysArgAsnLeuProValMetGlyAspGlyLeuGluThrSer 24
 Db 67 GCAGCACCGAGGAGGAGGATTCGAAACCCCTCTCTTGTATGAGGCTTTGAGAGCCCC 126
 QY 25 GlnMetSerThrThrGlnAlaGlnAlaGlnProGlnProAlaAsnAlaAlaSerThrAsn 44
 Db 127 ACAATGGCTTGGTGGCTGCT---TTGCAACTTACCCCTGCTC-----AAC 168
 QY 45 ProProProGluThrSerAsnProAsnLysProLysArgGluThrArgGluLeuGln 64
 Db 169 CCACCACCCCGGAGGTGTCCAAATCCAAAAAGCCAGAGAGATTACCAACACAGCTGCA 228

```

Oy 65 TyrlLeuAargValValLeuLysThrLeuTrpLysHisGlnPheAlaTrpProPheGln 84
Db 229 TACCTCGCGCAGGATGATGATGAAGGCTCTGTGGAAACATCATGCTCCCATGGCCATTCGG 288
Oy 85 GlnProValAspAlaValLysLeuAsnLeuProAsp----- 96
Db 289 CAGCCTGTGATGTCTGTCAAACTGGGCTATACGAT--TCCCACTCGGGTTGGAGAGAGAC 347
Oy 96 ----- 96
Db 348 CACGGTGGCCAAATCTTACGCTTCTTCCCTTCATGCAAGCCCATGATGACCAACC 407
Oy 97 -----TyTrpLysIleIleLysThrProMetAspMetGlyThrIleLysLysArg 113
Db 408 CCAGAGGATATATCACAAAATATTAATAACAGCCTATGAGACATGGGCTACTATTAGAGAGAGA 467
Oy 114 LeuGluAsnAsnTrpTyTrpTrpAsnAlaGlnGluCysIleGlnAspPheAsnThrMetPhe 133
Db 468 CTGGAAACAAATATTATTTAGGGGCTTCAGAGTGTATGCAAGATTTTAAATACCATGTC 527
Oy 134 ThrAsnCysTrpIleTyTrpAsnLysProGlyAspAspIleValLeuMetAlaGluAlaLeu 153
Db 528 ACCAATGTTACCTTATTCACAGCAAGCCCATGATGATATTTCTTAATGGCACAAGCCTG 587
Oy 154 GluLysLeuPheLeuGlnLysIleAsnGluLeuProThrGluThrGluIleMetIle 173
Db 588 GAAAGATATTCCTTACAGAAAGTTGCATCATGCTCAAGAAAGAACAGAGCTGTAGTG 647
Oy 174 ValGlnAlaLysGly-----ArgGly-----ArgGlyLys 184
Db 648 ACCATCCCTAAGAAACAGCAAGAAAGGGGCCAAGTTGGCAGCGCTCCAGGGCACTGTT 707
Oy 185 GluThrGlyThrAlaLysProGlyValSerThrValProAsnThrThrGlnAlaSerThr 204
Db 708 ACCAGTGGCCCATAGGTGCTGCTGCTCTCTGTGTGTACACACAGCCCTGTATCTCT 767
Oy 205 ProProGlnThrGlnThrProGlnProAsnProProProValGlnAlaThrProHisPro 224
Db 768 CCACCTGAGATA-----CTTACCACTGCTCTCAACATTTCCCAACCA 809
Oy 225 PheProAlaValThrProAspLeuIleValGlnThrProValMetThrValValProPro 244
Db 810 -----TCAGTCATTTTCTTCCACTTCTCAAGCTCTTGAC-- 844
Oy 245 GlnProLeuGlnThrProProValProProGlnProGlnProGlnProAlaProAla 264
Db 845 -----TCTGCTGAGCCCGCTCTCTGCTGTACTGACAGCTCTCCAGCC----- 889
Oy 265 ProGlnProValGlnSerHisProProIleIleAlaAlaThrProGlnProValLysThr 284
Db 890 -----CAGCCCTTGGCCAAG 904
Oy 285 LysLysGlyValLysArgLysAlaAspThrThrThrProThrThrIle----- 300
Db 905 AAAAAAGGCGT--AAGCGGAAGCAGATACTACCAACCCCTACACCTACAGCCATCTTGCT 963
Oy 301 -----AspProIleHisGlnProProSerLeuProProGlnProLysThrThrLysLeu 318
Db 964 CTTGTTCTTCCACTAGCTCTCTGGAGTCTT-----GAGCCTAAGGAGCAGACCCCTT 1017
Oy 319 ---GlyGlnArgArgLysSerSerArgProValLysProProLysLysAspValProAsp 337
Db 1018 CCCCCATGTGTAGAGAGATGCTGCCCCCATGCAAGCCCCCAGCAAAAGCTTGCTGAC 1077
Oy 338 SerGlnGlnHisProAlaProGluLysSerSerLysValSerGluGlnLeuLysCysCys 357
Db 1078 TCTCACACCAACACACAGAGCTTAAAGAAAG--AAAGCTTCAGAACAGTTAAACAATTGC 1136
Oy 358 SerGlyIleLeuLysGlnMetPheAlaLysLysHisAlaAlaLysTrpAlaTrpProPheTrp 377
Db 1137 AATGAGATT--TTGAAGGAGTTACTCTCTAAGAGCATGCTCTCTATGCTTCTTAT 1195
Oy 378 LysProValAspValGluAlaLeuGlyLeuHisAspTrpCysAspIleIleLysHisPro 397

```

```

Db 1196 AAACCGTGAATGCTTCTCTCACTTGCGCTGCATGACTTACCATGACATCATTAAGCCCC 1255
Oy 398 MetAspMetSerThrIleLysSerLysLeuGluAlaArgGluTrpArgAspAlaGlnGlu 417
Db 1256 ATGACACTCAGACACTCTCTCAAGCGGAAGATGAGAGAACCGTATTACCGGATGACAGAG 1315
Oy 418 PheGlyAlaAspValArgLeuMetPheSerAsnCysTrpLysTrpAsnProProAspHis 437
Db 1316 TTGCTGTGATGTACGGCTTATGTC--TCCAATGCTGTATAGTACAT--GCCCAAGATCAC 1373
Oy 438 GluValAlaAlaMetAlaArgLysLeuGlnAspValPheGluMetArgPheAlaLysMet 457
Db 1374 GATGTGTGGCAATGGCAGCAAGCTTACAGATGAT--TACTTCGTTATGCAAGATG 1430
Oy 458 ProAspGluProGluGluProValAlaValSerSerProAlaValProProPro----- 476
Db 1431 CCAGATGAACCACTAGAACCAAGGCTTATACAGTCTTACTGCTCCATGCCCCCTGCTG 1490
Oy 477 ThrLysValAlaAlaProProSerSerSerAspSerSerSerSerSer----- 494
Db 1491 GCCAATTCGTTCTTCAAGATGCTCCAGTGAAGAAAGTACGAGTGAAGCTCTCTGAGAA 1550
Oy 495 -----AspSerAspSerSerThrAspAspSerGlu 504
Db 1551 GAGAGAGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1610
Oy 505 GluGluArgAlaGlnArgLeuAlaGlnLeuGlnGlnLysAlaValHisGlnGln 524
Db 1611 GAGAAAGAGGCTCATGCTTATGACAGAACTTACAGAAACACTTCTGGGCACTGTAACAA 1670
Oy 525 LeuAlaAlaLeuSerGlnProGlnGlnAsnLysProLysLysGlnLysAspLysLys 544
Db 1671 CTGGCTGCTCTGTCCAGAGGTCCGATATCCAGGCCCAAGAGAGAGAGAGAGAGAGAGAG 1730

RESULT 9
ABV25740
ID ABV25740 standard; cDNA, 1862 BP.
XX
AC ABV25740;
XX
DT 16-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 25731.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN MO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US05171.
XX
PR 17-FEB-2000; 2000US-183319P.
PR 16-MAR-2000; 2000US-189862P.
PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
XX prostate cells and correlating with presence of prostate cancer, useful
XX for detecting presence of prostate cancer, stage of prostate cancer -
XX

```

PS Claim 1, Page 5167-5168, 11750dp; English.

XX The invention relates to an isolated nucleic acid molecule (1) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (1) is useful for:

- CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (1) is also useful as a pharmacodynamic or pharmacogenomic marker.

XX Sequence 1862 BP, 502 A, 514 C, 483 G, 363 T, 0 other;

Alignment Scores:

Pred. No.:	1,22e-47	Length:	1862
Score:	1161.50	Matches:	274
Percent Similarity:	59.40%	Conservative:	64
Best Local Similarity:	48.15%	Mismatches:	131
Query Match:	31.92%	Indels:	100
DB:	23	Gaps:	17

US-09-700-590A-22 (1-688) X ABV25740 (1-1862)

QY 126 ILEGLNAspPheAnthrMetPheThrAsnCysTyrIleTyrAnLysPProGlyAspasp 145
DB 16 ATTAATACTTGATACATCATGTTCCACCAACTGTGCTTACGACAAATATCATCATGAGGC 75
QY 146 ILEValLeuMetAlaGluAlaLeuGluLysLeuPheLysIleAsnGluLeuPro 165
DB 76 TTGTTGTAATGTAAACAAACGCTGGAAGATATTCCTACAGAAAGTTGCACTCAATGCCA 135
QY 166 ThcGluGluThrGluIleMetIleValGlnAlaLys----- 177
DB 136 CAAAGAAACAAAGAGCTGTGATGACCATCCCTAAGAACACAGACAGAGGGCCAAATT 195
QY 178 GlyArgGly-ArgGlyArgLysGluThrGlyThrAlaLysProGlyValSerThrValPr 197
DB 196 GCGAGCGCTCGGGGAGGTGTTACACAGTCCATCAGGTGCTCGCTCTCTCTGTC 255
QY 197 oAsnThrThrGlnAlaSerThrProProGlnThrGlnThrProGlnProAsnProPror 217
DB 256 ACACACAGCCCTATATCTCTCTCCACCTGAGATA----- 297
QY 217 oValGlnAlaThrProHisProPheProAlaValThrProAsnLeuIleValGlnThrPr 237
DB 298 TGTCCTCAACATTCCTCCACCA-----TCAGTCATTTCTCTCC 336
QY 237 oValMetThrValValProProGlnProLeuGlnThrProProValProProGlnPr 257
DB 337 ACTTCTCAAGCTCTTGAC-----TCGCTGAGACCCCGCTCTCTGTTAC 384
QY 257 oGlnProProAlaProAlaProGlnProValGlnSerHisProProIleIleAlaAl 277
DB 385 TGCAGCTCTCTCCAGC----- 400
QY 277 aThrProGlnProValLysThrLysGlyValLysArgLysAlaAspThrThrThrPr 297
DB 401 -----CAGCCCTTGCCAAAGAAAGGCGTTAAAGCGAAAGAGATCTACACACCC 453
QY 297 oThrThrIle-----AspProIleHisGluProProSerLeuProPr 311
DB 454 TACACCTACAGCACTTGGCTCGCTGTTCTCCACCTAGCCCTCTGGAGAGCTT----- 508
QY 311 oGluProLysThrThrLysLeu---GlyGlnArgArgLysSerSerArgProValLysPr 330
DB 509 -GAGCTTAAGGACAGACGCTTCCCTATGCGTAGAGAGAGTGTGCGCCCATCAAGCC 567

QY 330 oProLysLysAspValProAspSerGlnGlnHisProAlaProGluLysSerLysVal 350
DB 568 CCCACCCAAAGACTTCTGACCTCTGACACCAACACACACCTTAAAGAAAGAGCT 627
QY 350 lSerGluGlnLeuLysCysCysSerGlyIleLeuLysGluMetPheAlaLysLysHisAl 370
DB 628 TTCAAGACACTTAAACATTGCAATGCACTTTGAAGAGTTACTCTCAAGAGATCG 687
QY 370 aaAlaTyrAlaTyrProPheTyrLysProValAspValGluAlaLeuGlyLeuHisAspTy 390
DB 688 TGCCCTATGCTTGCCCTTTCTATTAACAGAGATGCTTCTGACCTTGAGCTGATGACTA 747
QY 390 rCysAspIleIleLysHisProMetAspMetSerThrIleLys----- 404
DB 748 CCATGACATCATTAACACCCCATGACCTCAGACCTGTCACTTCTGATGCTGCTCC 807
QY 405 -----Ser-LysLeuGluAlaArgGluTyrArgAspAlaGlnGluPheGlyAlaAspV 422
DB 808 TTCTGACGCGAAGATGAGAACCGTGATTAACCGGATGACAGAGCTTGTGCTGATG 867
QY 422 aAlaArgLeuMetPheSerAsnCysTyrLysTyrAsnProProAspHisGluValValAlaM 442
DB 868 TACGGCTTATGTTCTCCAACTGCTATAGTACATCCCGACATCAGATGTTGGCAA 927
QY 442 eAlaAlaArgLysLeuGlnAspValPheGluMetArgPheAlaLysMetProAspGluProG 462
DB 928 TGGCAGCAAAAGCTACAGAGATGATTTGAGTTCGTTATGCAAGATGCCAGATGAACAC 987
QY 462 lGluGluProValAlaAlaValSerSerProAlaValProProPro---ThrLysValValAl 481
DB 988 TAGAACCAAGGCGCTTTTACACGCTCTTACTGCTCATGCCCCCTGCTGGCCAAATCGCTT 1047
QY 481 lAproProSerSerSerAspSerSerSerSerSerSerSer----- 494
DB 1048 CAGAGTCTCTCAGTGAAGAAAGTAGCAGTGCAGCTCTCTGAGAAAGAGAGAGAG 1107
QY 495 -----AspSerAspSerSerThrAspAspSerGluGluAlaArgAlaG 509
DB 1108 ATGAGAGAGACGAGAGAGAAAGAGAGTGAAGCTCAGACCTCAGAGAGAAAGGCTC 1167
QY 509 lnaArgLeuAlaGluLeuGlnGluGlnLeuLysAlaValHisGluGlnLeuAlaLeuS 529
DB 1168 ATGCTTAAAGCAACTACAGAAACGCTTGGGAGATGATGAACAACTGGCTCTGT 1227
QY 529 eArgInProGlnAsnLysProLysLysLysLysLysAspLysLysGluLysLysLys- 548
DB 1228 CCCAGGCTCCAAATATCCAAAGCCCAAGAGAAA---AGAGAGAAAGAAAGAAAGAGA 1284
QY 549 -----GluLysHisLysArgLysGluGluValGluGluAsnLysLysSer---L 564
DB 1285 AACGGAAGCAGAGAAACATGAGAGCGAGCTGGGCGGATGAAGAGACAGAGGCTTA 1344
QY 564 ysaAlaLysGluProPro---ProLysLysThrLysLysAsnAsnSerSerAsnSerAnv 583
DB 1345 GGGACACCCCGCCCACTCAACCTAAGAGTCCAAAGAAAGCAAGTGGAGTGGGGTGGCA 1404
QY 583 alSerLysLysGluPro-----AlaProm 591
DB 1405 GTGCTGCTTAAGGCTCTTGGCTTGGACCTTGGAGGAAGTGGACCAAGCTCC 1464
QY 591 eLysSerLys-----Pro-ProProThrTyrGluSerGluGluGlu 604
DB 1465 AAAAAGCCCAAGAGACAGCCCACTGCTGCTCAAGATTATGATTACAGAGAGAG 1524
QY 605 AspLysCysLysPProMetSerTyrGluGluLysArgLysLeuSerLeuAspIleAsnLys 624
DB 1525 GAAAGAGACAGCCCAATGAGTTAGATGAAGGCGAGCTGAGCTGAGCATCAACAAA 1584
QY 625 LeuProGluGluLysLeuGly 631
DB 1585 TTACTGAGGCCCCAAGAGGA 1605

RESULT 10
 AAS03029
 ID AAS03029 standard; cDNA; 667 BP.
 XX
 XX AAS03029;
 AC
 XX
 DT 29-AUG-2001 (first entry)
 XX
 XX Human diagnostic and therapeutic (dthp) cDNA sequence #18.
 DE
 XX Human diagnostic and therapeutic molecule; dthp; gene therapy;
 KW thalassemia; cardiovascular disorder; cell proliferative disorder;
 KW cancer; neurodegenerative disorder; autoimmune disorder;
 KW infectious disorder; inflammatory disorder; developmental disorder;
 KW Incyte ID number 2432679dec; intracellular signalling molecule; ss.
 OS Homo sapiens.
 XX
 XX WO200121836-A2.
 PN
 XX
 PD 29-MAR-2001.
 XX
 PF 19-SEP-2000; 2000MO-US25643.
 XX
 XX 23-SEP-1999; 99US-0155760.
 PR 24-SEP-1999; 99US-0155939.
 PR 24-SEP-1999; 99US-0156294.
 PR 28-SEP-1999; 99US-0156565.
 PR 28-SEP-1999; 99US-0156624.
 PR 28-SEP-1999; 99US-0156625.
 PR 24-NOV-1999; 99US-0167410.
 PR 24-NOV-1999; 99US-0167453.
 PR 24-NOV-1999; 99US-0167517.
 PR 24-NOV-1999; 99US-0167520.
 PR 29-NOV-1999; 99US-0167542.
 PR 29-NOV-1999; 99US-0167943.
 PR 29-NOV-1999; 99US-0167945.
 PR 30-NOV-1999; 99US-0168197.
 PR 30-NOV-1999; 99US-0168265.
 PR 30-NOV-1999; 99US-0168429.
 PR 30-NOV-1999; 99US-0168432.
 PR 01-DEC-1999; 99US-0168468.
 PR 01-DEC-1999; 99US-0168599.
 XX
 XX (INCYTE) INCYTE GENOMICS INC.
 PA
 XX Hodgson DM, Lincoln SE, Russo FD, Spiro PA, Banville SC;
 PI Bratcher SR, Dufour GE, Cohen HU, Rosen BH, Shah P, Chalup MS;
 PI Hillman JL, Jones AL, Yu JY, Greenwalt LB, Panzer SR;
 PI Roseberry AM, Wright RJ, Chen W, Liu TF, Yap PE, Stockdreher TK;
 PI Amshey S, Fong WT;
 PI
 XX MPI; 2001-281607/29.
 DR
 XX
 XX Novel diagnostic and therapeutic polynucleotides, used in disease
 PT diagnosis and for gene therapy of conditions such as cancer and
 PT thalassemia
 PT
 XX
 XX Claim 1; Page 263; 299pp; English.
 PS
 XX
 XX The present sequence for human diagnostic and therapeutic (dthp) cDNA
 CC sequence #18 is 1 of 71 (AAS03012-AAS03082) novel sequences described
 CC in the invention. The present sequence (Incyte ID No: 2432679dec)
 CC encodes an intracellular signalling molecule. The dthp polynucleotides
 CC may be used to diagnose a condition disease or disorder associated with
 CC human molecules. They can be used to identify the presence of similar
 CC nucleic acids. Dthp polynucleotides may be used to generate hybridisation
 CC probes for use in chromosomal mapping. Polypeptides (DTHP) encoded by
 CC dthp are used to screen for molecules which bind to them and modulate
 CC their activity. Dthp polynucleotides can be used for gene therapy of
 CC disorders such as severe combined immunodeficiency syndrome (SCID),
 CC cystic fibrosis, thalassemia, haemophilia resulting from Factor VIII
 CC or IX deficiencies, cardiovascular disorders e.g familial

CC hypercholesterolaemia (RH), cell proliferative disorders e.g. cancers,
 CC neurodegenerative disorders, autoimmune/inflammatory disorders,
 CC infectious disorders and developmental disorders. The antibodies can be
 CC used to analyse protein expression levels.
 CC
 XX

SQ Sequence 667 BP; 229 A; 157 C; 149 G; 132 T; 0 other;

Alignment Scores:

Pred. No.:	2,23e-39	Length:	667
Score:	981.00	Matches:	186
Percent Similarity:	98.95%	Conservative:	3
Best Local Similarity:	97.38%	Mismatches:	1
Query Match:	26.96%	Indels:	1
DB:	22	Gaps:	0

US-09-700-590A-22 (1-668) x AAS03029 (1-667)

QY	1	MetSerAlaGluSerGlyPro-GlyThrArgLeuArgAsnLeuProValMetGlyAspG	20
DB	22	ATGTCGCGGAGAGCGCCCTGGGGAGAGATTGAGAAATCTGCCAGTATGGGGATG	81
QY	20	YLeuGluThrSerGluMetSerThrThrGlnArgGlnArgGlnProGlnProAlaAsnAl	40
DB	82	ACTAGAAACTTCCCAATGCTTAACAACAGAGCCAGGCCAACCCAGCCAGCAACG	141
QY	40	aAlaSerThrAsnProProProGluThrSerAsnProAsnLysProLysArgGlnTh	60
DB	142	AGCCACACCAACACCCCGCCGCCAGAGACCTCAACCTTAACAACCCAGAGGAGAC	201
QY	60	rAangGlnLeuGlnThrLeuLeuArgValValLeuLysThrLeuTrpLysHisGlnPheAl	80
DB	202	CAACCAACTGCATATACCTGCTCAGAGTGTGCTCAAGACACTATGAAACACCACTTTC	261
QY	80	aATPProPheGlnGlnProValAspAlaValLysLeuAsnLeuProAspTrpTrpLysI	100
DB	262	ATGGCTTTTCAGACAGCTGTGATCCGTCAGACTGAACTCTCCGATTACTATAGAT	321
QY	100	eLeuLysThrProMetAspMetGlyThrIleLysLysArgLeuGluAsnAsnTrpTr	120
DB	322	CATTAAACCCCTATGATGATGGAACATTAAGAAGCCCTTGAAACCAACATTTACG	381
QY	120	pAanaGlnGlnLysCysIleGlnAspPheAsnThrMetPheThrAsnCysTrpIleTrAs	140
DB	382	GAATGCTCAGGAATGATATCAGACTTCAACCTATGTTTACAAATGTTTACATCAAA	441
QY	140	nLysProGlyAspAspIleValLeuMetAlaGluAlaLeuGluLysLeuPheLeuGlnI	160
DB	442	CAAGCTTGAGATGACATGCTTAAATGACAGAGCTTGGAAGAACTCTTCTTGCAAAA	501
QY	160	sIleAsnGluLeuProThrGluGluThrGluIleMetIleValAlaGlnAlaLysGlyArgG	180
DB	502	AATTAATGAGCTTACCCACAGAGAAACCCAGATCATGATATGTCAGGCAAGAGAGAG	561
QY	180	yArgGlyArgLysGluThrGlyThrAlaLys 190	
DB	562	ACGTGGAGAAAAGAAAGTGGGTTATCAAG 592	

RESULT 11
 AAS26193
 ID AAS26193 standard; cDNA; 746 BP.
 XX
 XX AAS26193;
 AC
 XX
 DT 07-NOV-2001 (first entry)
 XX
 XX Human cDNA encoding a novel secreted protein, Seq ID 372.
 DE
 XX Human: immunosuppressive; antiarthritic; ss; antirheumatic;
 KW cytostatic; cardiac; vasotropic; cerebroprotective; nootropic;
 KW neuroprotective; antibacterial; varicide; fungicide; ophthalmological;
 KW vulnary; secreted protein; rheumatoid arthritis;
 KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
 KW cerebrovascular disorder; cerebral ischemia; angiogenesis;

KM nervous system disorder; Alzheimer's disease; infection; ocular disorder;
KM corneal infection; wound healing; epithelial cell proliferation;
KM skin aging; food additive; preservative; antiproliferative.

OS Homo sapiens.

PN WO200155322-A2.

PD 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US01341.

XX 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 14-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.

PR 26-JUL-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225256.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.

PR 14-AUG-2000; 2000US-0225759.

PR 18-AUG-2000; 2000US-0226279.

PR 22-AUG-2000; 2000US-0226681.

PR 22-AUG-2000; 2000US-0226868.

PR 23-AUG-2000; 2000US-0227182.

PR 21-SEP-2000; 2000US-0234274.

PR 25-SEP-2000; 2000US-0234997.

PR 25-SEP-2000; 2000US-0234998.

PR 26-SEP-2000; 2000US-0235484.

PR 27-SEP-2000; 2000US-0235834.

PR 27-SEP-2000; 2000US-0235836.

PR 29-SEP-2000; 2000US-0236327.

PR 29-SEP-2000; 2000US-0236367.

PR 29-SEP-2000; 2000US-0236368.

PR 29-SEP-2000; 2000US-0236369.

PR 29-SEP-2000; 2000US-0236370.

PR 02-OCT-2000; 2000US-0236802.

PR 02-OCT-2000; 2000US-0237037.

PR 02-OCT-2000; 2000US-0237038.

PR 02-OCT-2000; 2000US-0237039.

PR 02-OCT-2000; 2000US-0237040.

PR 13-OCT-2000; 2000US-0239935.

PR 13-OCT-2000; 2000US-0239937.

PR 20-OCT-2000; 2000US-0240960.

PR 20-OCT-2000; 2000US-0241221.

PR 20-OCT-2000; 2000US-0241785.

PR 20-OCT-2000; 2000US-0241785.

PR 20-OCT-2000; 2000US-0241787.

PR 20-OCT-2000; 2000US-0241808.

PR 20-OCT-2000; 2000US-0241809.

PR 20-OCT-2000; 2000US-0241826.

PR 01-NOV-2000; 2000US-0244617.

PR 08-NOV-2000; 2000US-0246474.

PR 08-NOV-2000; 2000US-0246475.

PR 08-NOV-2000; 2000US-0246476.

PR 08-NOV-2000; 2000US-0246477.

PR 08-NOV-2000; 2000US-0246478.

PR 08-NOV-2000; 2000US-0246523.

PR 08-NOV-2000; 2000US-0246524.

PR 08-NOV-2000; 2000US-0246525.

PR 08-NOV-2000; 2000US-0246526.

PR 08-NOV-2000; 2000US-0246527.

PR 08-NOV-2000; 2000US-0246528.

PR 08-NOV-2000; 2000US-0246532.

PR 08-NOV-2000; 2000US-0246609.

PR 08-NOV-2000; 2000US-0246610.

PR 08-NOV-2000; 2000US-0246611.

PR 08-NOV-2000; 2000US-0246613.

PR 17-NOV-2000; 2000US-0249207.

PR 17-NOV-2000; 2000US-0249208.

PR 17-NOV-2000; 2000US-0249209.

PR 17-NOV-2000; 2000US-0249210.

PR 17-NOV-2000; 2000US-0249211.

PR 17-NOV-2000; 2000US-0249212.

PR 17-NOV-2000; 2000US-0249213.

PR 17-NOV-2000; 2000US-0249214.

PR 17-NOV-2000; 2000US-0249215.

PR 17-NOV-2000; 2000US-0249216.

PR 17-NOV-2000; 2000US-0249217.

PR 17-NOV-2000; 2000US-0249218.

PR 17-NOV-2000; 2000US-0249244.

PR 17-NOV-2000; 2000US-0249245.

PR 17-NOV-2000; 2000US-0249264.

PR 17-NOV-2000; 2000US-0249265.

PR 17-NOV-2000; 2000US-0249297.

PR 17-NOV-2000; 2000US-0249299.

PR 17-NOV-2000; 2000US-0249300.

PR 01-DEC-2000; 2000US-0250160.

PR 01-DEC-2000; 2000US-0250391.

PR 05-DEC-2000; 2000US-0251030.

PR 05-DEC-2000; 2000US-0251988.

PR 06-DEC-2000; 2000US-0256719.

PR 08-DEC-2000; 2000US-0251856.

PR 08-DEC-2000; 2000US-0251868.

PR 08-DEC-2000; 2000US-0251869.

PR 08-DEC-2000; 2000US-0251989.

PR 08-DEC-2000; 2000US-0251990.

PR 21-SEP-2000; 2000US-0234223.

PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 DR WPI; 2001-488783/53.
 XX P-PSDB; AAU16206.

PT New nucleic acid molecules encoding 461 human secreted proteins for
 PT diagnosing, preventing, treating or ameliorating medical conditions and
 PT used as food additives or preservatives -

PS
 XX Claim 1; SEQ ID No 372; 980bp; English.

CC The invention relates to isolated nucleic acid molecules and their
 CC encoded secreted proteins. The nucleic acids and proteins are used to
 CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They
 CC are also used in diagnosing a pathological condition or susceptibility
 CC to a pathological condition. Antibodies to the proteins can also
 CC be used in alleviating symptoms associated with the disorders and in
 CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
 CC immunosorbent assays (ELISA). Disorders which are diagnosed or treated
 CC include autoimmune diseases e.g. rheumatoid arthritis,
 CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
 CC e.g. cerebral ischemia, angiogenesis, nervous system disorders e.g.
 CC Alzheimer's disease, infections caused by bacteria, viruses and fungi
 CC and ocular disorders e.g. corneal infection, and many other
 CC disorders listed in the specification. The polypeptides can also
 CC be used to aid wound healing and epithelial cell proliferation, to
 CC prevent skin aging due to sunburn, to maintain organs before
 CC transplantation, for supporting cell culture of primary tissues, to
 CC regenerate tissues and in chemotaxis. The polypeptides can also be used
 CC as a food additive or preservative to increase or decrease storage
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
 CC minerals, cofactors and other nutritional components. The present
 CC sequence encodes a novel secreted protein of the invention.

Alignment Scores:

Pred. No.: 3,05e-25 Length: 746
 Score: 688.00 Matches: 144
 Percent Similarity: 66.40% Conservative: 22
 Best Local Similarity: 57.60% Mismatches: 61
 Query Match: 18.91% Indels: 24
 DB: 22 Gaps: 6

US-09-700-590A-22 (1-688) x AAS26193 (1-746)

QY 51 SerAsnProAsnlySProlYsArgGlnThraSglnLeuGlnTYrLeuLeuArgValVal 70
 DB 3 TCCAAACCCGAGCAAGCCGCGCCGCAAGCAACCAAGTGTGAGTGTG 62
 QY 71 LeuYsThrLeuTyrPlySHISglnPheAlaTyrProPheGlnInProValAspAlaVal 90
 DB 63 GTGAAGACGCTCTGGAACACCAAGTTCGCTGCGCTTCTTCAAGCCGCGGAGCAATC 122
 QY 91 LysLeuAsnLeuProAspTYrTYrLysIleIleLysThrProMetAspMetGlyThrIle 110
 DB 123 AAATTGAACCTGCGCGGATTCATCAAAATATATTAACCAACCAATGATGGGACATTT 182
 QY 111 LysLysArgLeuGlnuAsnAsnTYrTYrTyrAsnAlaGlnGluCysIleGlnAspPheAsn 130
 DB 183 AAGAGAGACTAGAAATAATATTATTATTTGAGTCCAGACCGAATGTATGAGACTTCAAC 242
 QY 131 ThrMetPheThrAsnCysTYrIleTyrAsnLysProGlyAspAspIleValLeuMetAla 150
 DB 243 ACCATGTTTCAAAATGTTACATTATTAACAGCCCAAGTGCATATGAGCTATGACC 302
 QY 151 GluAlaLeuGlnLysLeuPheLeuGlnLysIleAsnGluLeuProThrGlnGluThrGlu 170

DB 303 CAAGCTTTAGAGAAATTTTCTTACAAAAAGTGCCAGATGCCCAAGAGAAGTTGA 362
 QY 171 IleMetIleValGlnAlaLysGlyArgGlyArgGlyArgGlyArgGlyThrAlaLys 190
 DB 363 ATTATTATACCCCTGCTCCAAAGAGGNCMAAGTGAGAGCCGCT-GGGGAGCCAGAGC 421
 QY 191 ProGlyValSerThrValProAsnThrThrGlnAlaSerThrProGlnThrGlnThr 210
 DB 422 GCAGGTACACAGCAAGTGGGCGCGCTCTGTCTCTCC-----CCAGCGACC 469
 QY 211 Pro---GlnProAsnProProProValGlnAlaThrProHisProPheProAlaValThr 229
 DB 470 CCTTTAGAGCGTCCCGCCACCGTCTCCAGAGCC----- 508
 QY 230 ProAspLeuIleValGlnThrProValMetThrValValProGlnProLeuGlnThr 249
 DB 509 -----GTATCGCTGCGACCCCTGTACCAACATCATCTGCAAAC-----GTCACG 553
 QY 250 ProProProValProProGlnProGlnProProProAlaProAlaProGlnProValGln 269
 DB 554 TCGGTCCAGTCCCGCCAGCTGCGCCCACTCT-----CCTGCCACA 598
 QY 270 SerHisProProIleIleAlaAlaThrProGlnProValLysThrLysGlyValLys 289
 DB 599 CCATCGCTCCCGGTGGTCCCTCTAGCCCA--CTGTCTCAAGAAANAGCGTTGAA 655
 QY 290 ArgLysAlaAspThrThrThrProThrThr 299
 DB 656 CGGAAGCAACACACACACTCCACGACT 685
 RESULT 12
 AAS26606
 ID AAS26606 standard; cDNA; 424 BP.
 XX
 AC AAS26606;
 XX
 DT 07-NOV-2001 (first entry)
 XX
 DE Human cDNA encoding a novel secreted protein, Seq ID 785.
 XX
 KW Human; immunosuppressive; antiarthritic; ss; antineumatic;
 KW cytoskeletal; cardiac; vasotrophic; cerebroprotective; nootropic;
 KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
 KW vulnary; secreted protein; rheumatoid arthritis;
 KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
 KW cerebrovascular disorder; cerebral ischemia; angiogenesis;
 KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;
 KW corneal infection; wound healing; epithelial cell proliferation;
 KW skin aging; food additive; preservative; antiproliferative.
 OS Homo sapiens.
 XX
 PF 17-JAN-2001; 2001MO-US01341.
 XX
 XX 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-020515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 14-AUG-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226688.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236137.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0237040.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.

PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 17-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249266.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 06-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 (HUMA-) HUMAN GENOME SCI INC.
 Rosen CA, Barash SC, Ruben SM;
 WPI, 2001-488763/53.
 DR P-PSDB; AAU16619.
 XX
 PT New nucleic acid molecules encoding 461 human secreted proteins for
 PT diagnosing, preventing, treating or ameliorating medical conditions and
 PT used as food additives or preservatives -
 PT
 PS Claim 1; SEQ ID No 785; 980bp; English.
 XX
 CC The invention relates to isolated nucleic acid molecules and their
 CC encoded secreted proteins. The nucleic acids and proteins are used to
 CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They
 CC are also used in diagnosing a pathological condition or susceptibility
 CC to a pathological condition. Antibodies to the proteins can also
 CC be used in alleviating symptoms associated with the disorders and in
 CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
 CC immunoassay assays (ELISA). Disorders which are diagnosed or treated
 CC include autoimmune diseases e.g. rheumatoid arthritis,
 CC hyperproliferative disorders e.g. neoplasms of the breast or liver,

CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
 CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
 CC Alzheimer's disease, infections caused by bacteria, viruses and fungi
 CC and ocular disorders e.g. corneal infection, and many other
 CC disorders listed in the specification. The polypeptides can also
 CC be used to aid wound healing and epithelial cell proliferation, to
 CC prevent skin aging due to sunburn, to maintain organs before
 CC transplantation, for supporting cell culture of primary tissues, to
 CC regenerate tissues and in chemotaxis. The polypeptides can also be used
 CC as a food additive or preservative to increase or decrease storage
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
 CC minerals, cofactors and other nutritional components. The present
 CC sequence encodes a novel secreted protein of the invention.

Alignment Scores:

Pred. No.:	1,65e-19	Length:	424
Score:	564.00	Matches:	105
Percent Similarity:	85.71%	Conservative:	15
Best Local Similarity:	75.00%	Mismatches:	19
Query Match:	15.50%	Indels:	2
DB:	22	Gaps:	1

US-09-700-590A-22 (1-688) x AAS26606 (1-424)

QY 51 SerAsnProAsnLysProLysArgGlnThrAsnGlnLeuGlnIleuLeuArgValVal 70
 |||||
 DB 3 TCCACCCCGAGCAAGCCCGCCGAGACCAACAGCTGAGTACAGCAAGTGTGTG 62
 |||||
 QY 71 LeuLysThrIleuTrpLysHisGlnPheAlaTyrProPheGlnGlnProValAspAlaVal 90
 |||||
 DB 63 GTGAAAGCGCTGTGAAACACAGTTGCTGCTGCTTCTACAGCCCGTGAGCGCATC 122
 |||||
 QY 91 LysLeuAsnLeuProAspTyrTyrLysIleIleLysThrProMetAspMetGlyThrIle 110
 |||||
 DB 123 AAATTGAACCTGCCGATTTTCATTAATAATTAATAAACCAGATGATATGGGACTATT 162
 |||||
 QY 111 LysLysArgLeuGlnLysLeuAsnTyrTyrTyrAsnAlaGlnGlnCysIleGlnAspPheAsn 130
 |||||
 DB 183 AAGAGAGACTAGAAATATATTATTATGAGTGCAGCAAGTATATGAGGACTTCAAC 242
 |||||
 QY 131 ThrMetPheThrAsnCysTyrIleTyrAsnLysProGlyAspAspIleValIleMetAla 150
 |||||
 DB 243 ACCATGTTTCAAAATTTTCAATTATATACAGCCCAAGATGACATATGCTAATGGCC 302
 |||||
 QY 151 GlnAlaLeuGlnLysLeuPheLeuGlnLysIleAsnGlnLeuProThrGlnGlnThrGlu 170
 |||||
 DB 303 CAAGCTTTAGAGAAATTTTCTACAAAAGTGGCCAGATGCCCAAGAGAGAGTTGAA 362
 |||||
 QY 171 IleMetIleValGlnAlaLysGlyArgGlyArgGlyArgLysGlnThrGlyThrAlaLys 190
 |||||
 DB 363 ATT---ATTACCCCTGCTCCAA-AGGGGCAAAAGTGTGGAAGCCGCTGGGGAGCCAG 418
 |||||

RESULT 13

AAZ80432
ID AAZ80432 standard; cDNA; 450 BP.

AC AAZ80432;

DT 07-APR-2000 (first entry)

DE Human colon cancer cell line SW480 cDNA clone SEQ ID NO:516.

XX Human: gene expression product; diagnosis; tumour; colon cancer;
 KM colorectal adenocarcinoma; cell line SW480; cell proliferation;
 KM cytotoxic; sarcoma; breast cancer; neoplasia; dysplasia;
 KM hyperplasia; de.

OS Homo sapiens.

XX MO9964576-A2.

XX 16-DEC-1999.

XX

PF 09-JUN-1999; 99WO-1B01062.

XX 10-JUN-1998; 98US-0088801.

XX (FARB) BAYER CORP.

XX Endege WO, Steimann KE, Astle JH, Burgess CC, Bushnell SE;

PI Carroll E, Catino TJ, Dertl A, Ford DM, Lewis ME, Monahan JE;

PI Schlegel R;

XX MPI, 2000-087220/07.

XX Novel nucleic acids, used to develop products for the diagnosis and

XX treatment of disorders involving unwanted cell proliferation,

XX particularly cancers, especially colon cancer

XX Claim 15; Page 336; 469pp; English.

XX AAZ79917 to AAZ80766 represent double stranded cDNA clones isolated from

XX the human colorectal adenocarcinoma (colon cancer) cell line SW480. The

XX cDNA clones can be used to generate antisense oligonucleotides which

XX can be used for antisense therapy. Methods and products from the present

XX invention can be used for identifying and/or classifying cancerous cells

XX present in a human tumour, particularly in solid tumours, e.g.

XX carcinomas and sarcomas, e.g. breast or colon cancers. The cDNA clones

XX can be used for developing agents for the diagnosis and treatment of

XX disorders involving unwanted cell proliferation, such as neoplasia,

XX dysplasia or hyperplasia.

XX Sequence 450 BP; 122 A; 127 C; 104 G; 96 T; 1 other;

XX Alignment Scores:

Pred. No.:	1.16e-15	Length:	450
Score:	484.50	Matches:	96
Percent Similarity:	75.00%	Conservative:	12
Best Local Similarity:	66.67%	Mismatches:	27
Query Match:	13.31%	Indels:	9
DB:	21	Gaps:	3

US-09-700-590A-22 (1-688) x AAZ80432 (1-450)

QY 286 LysGlyValLysArgLysAlaAspThrThrThrProThrThrIle----- 300
 |||||
 DB 3 AAAGCGTAAAGCGAAGAGAGATCTACACCCCTACACCTACACCTTGGCTCT 62
 |||||
 QY 301 ---AspProIleHisGlnProProSerLeuProProGlnProLysThrThrLysLeu--- 318
 |||||
 DB 63 GGTCTCCAGCTAGCCCTCTGGAGTCTT-----GAGCCTTAAGGACAGCAGGCTTCC 116
 |||||
 QY 319 GlyLysArgGlnLysSerSerArgProValLysProProLysLysAspValProAspSer 338
 |||||
 DB 117 CCTATGCGTAGAGAGAGTGTGCTGCCCATCATACAGCCCAAGCAAGATCTGACTCT 176
 |||||
 QY 339 GlnGlnHisProAlaProGlnLysSerSerLysValSerGlnGlnLysCysCysSer 358
 |||||
 DB 177 CAGCAACAACACAGAGCTCTAAGAAAGAGCTTTCAGAAACAGTTAAACATTCGAAT 236
 |||||
 QY 359 GlyIleLeuLysGlnMetPheAlaLysLysHisAlaIleTyrAlaTrpProPheTyrLys 378
 |||||
 DB 237 GGCATTTTGAAGAGTACTCTCTTAAGAGCATGCTGCTATGCTTGGCTTCTATAAA 296
 |||||
 QY 379 ProValAspValGlnAlaLeuGlnLysLeuHisAspTyrCysAspIleIleLysHisPomet 398
 |||||
 DB 297 CCAATGAGATGCTTCTGCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 356
 |||||
 QY 399 AspMetSerThrIleLysSerLysLeuGlnAlaArgGlnLysArgAspAlaGlnGlnLys 418
 |||||
 DB 357 GACCTCAGACGTCTCAAGCGGAAGATGAGAACCGTGAATTACCGGATGACAGAGTTT 416
 |||||
 QY 419 GlyAlaAspVal 422
 |||||
 DB 417 GCTGCTGATGTA 428
 |||||

XX	AB060244/C	standard; cDNA; 560 BP.
XX	AB060244;	
XX	02-AUG-2002	(first entry)
XX	Human colon cancer	related nucleotide sequence SEQ ID NO:3939.
XX	Human; colon cancer;	cancer; tissue profiling; forensic; mapping;
XX	genetic analysis;	diagnostic; antisense therapy; gene; ss.
XX	Homo sapiens.	
XX	MO200229086-A2.	
XX	11-APR-2002.	
XX	02-OCT-2001;	2001WO-US30732.
XX	02-OCT-2000;	2000US-237271P.
XX	(FARB)	BAYER CORP.
XX	Burgess C, Astle JH,	Carroll E, Catino TJ, Dwivedi P, Molino GA;
XX	Thiagalingam A, Lewis ME;	
XX	WPI; 2002-426115/45.	
XX	New isolated nucleic acid	that is differentially expressed in cancer
XX	tissues useful for	determining the presence of colon cancer in a cell
XX	or tissue type, and in	antisense therapy
XX	Claim 1; Fig 1;	796pp; English.
XX	AB060244	represent isolated nucleic acids (I) differentially
XX	expressed in cancer tissues.	AB060244 represent proteins
XX	encoded by the AB060244	nucleic acid sequences. (I) can be
XX	used in antisense therapy.	An antibody immunoreactive with a polypeptide
XX	encoded by (I) is useful	for detecting cancer in a patient sample, and
XX	for detecting the presence	or absence of a polynucleotide encoded by a
XX	nucleic acid which	hybridises to (I) in a cell. A probe/primer derived
XX	from (I) can be used	for determining the presence of a nucleic acid which
XX	hybridises to (I), and	for determining the phenotype of cells in a sample
XX	of cells from a patient.	(I) is useful for determining the presence of
XX	colon cancer in a cell	or tissue type, for determining the presence or
XX	state of other type of	cancer, in antisense therapy, to generate
XX	macroarrays on a solid	surface, to identify a chromosome on which the
XX	corresponding gene	resides, and in tissue profiling, forensics, genetic
XX	analysis, mapping and	diagnostic applications. (I) can be used to raise
XX	antibodies, and to	screen for peptide analogues and antagonists.
XX	Sequence 560 BP,	135 A, 121 C, 152 G, 141 T, 11 other;
XX	Alignment Scores:	
XX	Pred. No.:	2.99e-14 Length: 560
XX	Score:	457.00 Matches: 92
XX	Percent Similarity:	69.74% Conservative: 13
XX	Best Local Similarity:	59.35% Mismatches: 28
XX	Query Match:	12.56% Indels: 22
XX	DB:	24 Gaps: 3
XX	US-09-700-590A-22	(1-688) x AB060244 (1-560)
XX	281	ProValLysThrLysLeuGlyValIysArgLysAlaSerThr-----ThrThrPro 297
XX	443	CGTTGCGCAAGAAAGGCGTAAAGGAAAGCAATANTTTACCCACCCCTTACACCT 384
XX	298	ThrThrLysProIleHisGluProProSer----- 308
XX	383	ACAGCCATCTTGCGCTTGGTCTTCCAGCTAAGCCCTCTTGGAGAGCTTGAAGCATTAAAGCA 324

```

QY 309 -----leuProProGluInProLySThrThriLySLeuGlyGlnArgAGAGuSerSerArg 366
Db 323 GCACGGCTTCCCCCT-----ATGCTGAGAGAGAGCTGCTGC 288
QY 327 ProValLySProProLySAspValProAspSerGlnGlnHisProAlaProGluLyS 366
Db 287 CCCATCAAGCCCCCAGCAAGCAAGACTTGGCTGTGACTCAGCAACAACACAGAGCTTTAAG 228
QY 347 SerSerLySValSerGlnGlnLeuLySCysSerSerGlyIleLeuLySGluMetPheAla 366
Db 227 AAGGAAGAAGCTTTGAGAACACTTTAAACATTGCATATGGCATTTTGAAGAGATTACTCT 166
QY 367 LysLySHisAlaAlaIATyrAlaTyrProPheTyrLySProValAspValGluAlaLeuGly 386
Db 167 AAGAGCAATGCTGCTTAAGCTTGGCTTTCTATTAACACAGAGATGCTTCTGCACTTGGC 108
QY 387 LeuHisAspTyrCysAspIleIleLySHisProMetAspMetSerThrIleLySerLyS 406
Db 107 CTGCATGACTACATGACATCATTTAAGACACCCCATGGACCTCACACTGTCTCAACGGGAG 48
QY 407 LeuGlnAlaArgGluTyrArgAspAlaGlnGluPheGlyAlaArg 421
Db 47 ATGGAGAACCGGTGAATTACCGGAGATGCACAGAGATTTCCTGTGAT 3

```

RESULT 15
 ID AAX29128 standard; cDNA; 642 BP.
 AC AAX29128;
 DT 04-JUN-1999 (first entry)
 DE Polynucleotide RTP483 having hypoxia regulated activity.
 KW Hypoxia; angiogenesis; apoptosis regulated activity; ischemia; tumour;
 KW hypoxic retina; ss.
 OS Rattus sp.
 PN MO9909046-A1.
 PD 25-FEB-1999.
 PF 21-AUG-1998; 98MO-US17297.
 PR 21-AUG-1997; 97US-0056453.
 PA (KOHN/) KOHN K I.
 PA (QIAR-) QIAR BIOTECH INC.
 PI Einat P, Skalter R;
 DR WPI, 1999-180964/15.
 PT New polynucleotides with hypoxia, angiogenesis and apoptosis
 PT regulated activity - useful for diagnosis of ischemia, and for
 PT treatment involving gene therapy
 PS Claim 1; Page 55-56; 72pp; English.
 CC Sequences AAX29126 to AAX29136 represent isolated polynucleotides with
 CC hypoxia, angiogenesis and/or apoptosis regulated activity. The invention
 CC provides a method for diagnosing ischemia from bodily fluid samples by
 CC identifying at least one expressed gene, or identifying at least one up-
 CC regulated gene in a tissue sample. In situ analysis was performed for
 CC the above sequences in solid tumours and hypoxic retinas. The expression
 CC of the genes is activated in tumours around the hypoxic region in the
 CC tumour centre, therefore are hypoxia-regulated in vivo. Up regulation
 CC promotes angiogenesis required for tumour growth. The genes, proteins,
 CC antagonists, antisense oligonucleotides and dominant negative peptides
 CC are useful for regulating angiogenesis, apoptosis and hypoxia.
 CC Sequence 642 BP; 169 A; 180 C; 140 G; 151 T; 2 other;
 CC

Alignment Scores:

Pred. No.:	6,4e-13	Length:	642
Score:	430.50	Matches:	99
Percent Similarity:	59.41%	Conservative:	21
Best Local Similarity:	49.01%	Mismatches:	40
Query Match:	11.83%	Indels:	44
DB:	20	Gaps:	3

US-09-700-590a-22 (1-688) x AAX29128 (1-642)

```

QY      66  LeuleuArgValValLeuLysThrLeuTyrPLeuHisGlnPheAlaTrpProPheGlnGln 85
      |||  ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      3  CTGCACAGAGCTAGTAATGAGAGGCTCTGTGAGAGCATGATTGCATGCGCATTCGGCAG 62
QY      86  ProValAspAlaValLysLeuAsnLeuProAsp----- 96
      |||  |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      63  CCTGTGAGCGCTGTGAGAGCTGTGCGCAT-TCCACCCCTGTGGTGAGAGACCAC 121
QY      96  ----- 96
DB      122  GGTGGCCAAATTTCTTAGCTTCTTCTTCCTCATGACGCCATGATAGCCATCCCA 181
QY      97  -----TyrTyrLysIleIleLysThrProMetAspMetGlyThrIleLysArgLeu 114
      |||  ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      182  GAGGATTATCAAAATTTATCAACAGCCCATGAGCATGGAACTATCAAGAGAGACTT 241
QY      115  GluAsnAsnTyrTyrTrpAsnAlaGlnGluCysIleGlnAspPheAsnThrMetPheThr 134
      |||  |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      242  GAAAAACAATTACTAGCTGGCTGCTCAGAAATGTATGCAGATTTTAAACCACATGTTACC 301
QY      135  AsnCysTyrIleTyrAsnLysProGlyAspAspIleValLeuMetAlaGluAlaLeuGlu 154
      |||  |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      302  AACTGTTATTTTACAAACAGCCCATCGATGATTTGTCTTAAGSCACAGACACTGGA 361
QY      155  LysLeuPheLeuGln-LysIleAsnGluLeuProThrGluGluIleMetIleVal 174
      |||  ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      362  AAGATCTTCTTACGAAAGAGTCATGATGCCACAGAGAGCAGAGAGCTGTGTGAC 421
QY      174  IGlAlAlaLysGly-----ArgGly-----ArgGlyArgGlySGI 185
      |||  |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      422  TATCCCTTAAGAACAGCATAGAGAGGGGCCAAGTTAGACAGCACTCCAGGGCAGATTAC 481
QY      185  utnrglyThrAlaLysProGlyValSerThrValProAsnThrThrGlnAlaSerThrPr 205
      |||  |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      482  CAGTGGCCATCAGGTGCTGTCTGTCTTCTGTGTGCGATACAGCCCT-GTATANCCACC 540
QY      205  oProGlnThrGlnThr-ProGlnProAsnProProProValGlnAlaThrProHisProp 225
      |||  |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      541  ACCTGAATAACCTACGACCGCTCAACATTCACACCCCATCAGTCATCTGTCTNCCCT 600
QY      225  he 225
      |||
DB      601  TC 602

```

Search completed: March 14, 2003, 20:54:36
 Job time : 413 secs

THIS PAGE BLANK (uspto)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 14, 2003, 18:49:30 ; Search time 148 Seconds
(without alignments)
3262.724 Million cell updates/sec

Title: US-09-700-590a-22

Perfect score: 3639
Sequence: 1 MSASGPGTRLRNLPVMDG.....ALCHLLFAEKETFKLRKLM 688

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 501302 seqs, 350932545 residues

Total number of hits satisfying chosen parameters: 1002604

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cg2_1/USPTO_spool/US9700590/rnatat_14032003_140631_3291/app.query.fasta_1.839
-DB=Published Applications NA -QFMT=fastcap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=biosum62
-TRANS=human40.cdi -LIST=45 -DOCLALIGN=200 -THR SCORE=pcpt -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=plc -NORM=ext -HEADSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09700590 @CGN 1.1 80 @rnatat_14032003_140631_3291
-NCPU=6 -ICPU=3 -NO.XLPEXY -NO.NMAP -LARGEQUERY -NEG SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6
-Fgapext=7 -Ygapop=10 -Ygapext=0.5 -DELOP=6 -DELEX=7

Database : Published Applications NA:

1: /cg2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cg2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cg2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cg2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cg2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cg2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cg2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cg2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cg2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cg2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cg2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cg2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cg2_6/ptodata/2/pubpna/US6_NEW_PUB.seq:*
14: /cg2_6/ptodata/2/pubpna/US6_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1888	51.9	4664	10	US-09-880-107-1657 Sequence 1657, App
2	688	18.9	746	10	US-09-764-864-372 Sequence 372, App
3	564	15.5	424	10	US-09-764-864-785 Sequence 785, App
4	484.5	13.3	450	10	US-09-879-536-516 Sequence 516, App

5	362.5	10.0	429	10	US-09-876-889-243 Sequence 243, App
6	323.5	8.9	7326	12	US-10-109-886-7 Sequence 7, App1
7	308.5	8.5	547	10	US-09-815-343-213 Sequence 213, App
8	308	8.5	8147	12	US-10-109-886-9 Sequence 9, App1
9	272.5	7.5	358	10	US-09-783-590-1966 Sequence 1966, App
10	262.5	7.2	1855	10	US-09-864-761-19708 Sequence 19708, A
11	262.5	7.2	1958	10	US-09-864-761-2927 Sequence 2927, Ap
12	236	6.5	5037	9	US-09-808-880-1 Sequence 1, App1
13	226	6.2	1140	9	US-09-938-842A-1265 Sequence 1265, Ap
14	226	6.2	1140	9	US-09-887-576-652 Sequence 652, App
15	225	6.2	2653	10	US-09-764-864-373 Sequence 373, App
16	221.5	6.1	2825	10	US-09-822-889A-258 Sequence 258, App
17	221	6.1	2832	10	US-09-764-864-371 Sequence 371, App
18	220	6.0	2138	10	US-09-764-864-784 Sequence 784, App
19	220	6.0	2283	9	US-09-938-842A-1691 Sequence 1691, Ap
20	220	6.0	1989	9	US-09-887-576-645 Sequence 645, App
21	218.5	6.0	1998	9	US-09-894-844-6 Sequence 6, App1
22	218.5	6.0	2050	10	US-09-791-171-69 Sequence 69, App1
23	217.5	6.0	2824	12	US-10-124-557-13 Sequence 13, App1
24	217.5	6.0	3066	12	US-10-124-557-83 Sequence 83, App1
25	217.5	6.0	3117	12	US-10-124-557-73 Sequence 73, App1
26	217.5	6.0	3148	12	US-10-124-557-57 Sequence 57, App1
27	217.5	6.0	3420	12	US-10-124-557-103 Sequence 103, App
28	217.5	6.0	3813	12	US-10-124-557-43 Sequence 43, App1
29	217.5	6.0	3936	12	US-10-124-557-41 Sequence 41, App1
30	217.5	6.0	3942	12	US-10-124-557-141 Sequence 141, App
31	217.5	6.0	3945	12	US-10-124-557-49 Sequence 49, App1
32	217.5	6.0	3963	12	US-10-124-557-45 Sequence 45, App1
33	217.5	6.0	3963	12	US-10-124-557-59 Sequence 59, App1
34	217.5	6.0	4065	12	US-10-124-557-47 Sequence 47, App1
35	217.5	6.0	4086	12	US-10-124-557-39 Sequence 39, App1
36	217.5	6.0	4092	12	US-10-124-557-51 Sequence 51, App1
37	217.5	6.0	4215	12	US-10-124-557-61 Sequence 61, App1
38	217.5	6.0	4575	12	US-10-044-090-303 Sequence 303, App
39	217.5	6.0	5008	12	US-10-124-557-1 Sequence 1, App1
40	217.5	6.0	15720	9	US-10-025-380-1058 Sequence 1058, Ap
41	217.5	6.0	15720	10	US-09-922-217-1058 Sequence 1058, Ap
42	217.5	6.0	15720	10	US-09-833-263-1058 Sequence 1058, Ap
43	215.5	5.9	2272	9	US-10-174-590-345 Sequence 345, App
44	215.5	5.9	2272	9	US-10-176-758-345 Sequence 345, App
45	215.5	5.9	2272	9	US-10-175-737-345 Sequence 345, App

ALIGNMENTS

RESULT 1
US-09-880-107-1657
; Sequence 1657, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1657
; LENGTH: 4664
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 D42040
US-09-880-107-1657

Alignment Scores:

Pred. No.: 2.05e-104 Length: 4664
 Score: 1888.00 Matches: 801
 Percent Similarity: 67.08% Conservative: 81
 Best Local Similarity: 55.88% Mismatches: 142
 Query Match: 51.88% Indels: 97
 DB: 10 Gaps: 20

US-09-700-590a-22 (1-688) x US-09-880-107-1657 (1-4664)

QY 5 SerGlyProGluThrArgLeuArgAsnLeuProValMetGlyAspGlyLeuGluThrSer 24
 DB 1780 GCACACACGAGGAGAAAGATTGGAACCCCTCTCTTGTATGAGGCTTGAGAGCCCC 1839
 QY 25 GlnMetSerThrThrGlnAlaGlnAlaGlnProGlnProAlaAsnAlaAlaSerThrAsn 44
 DB 1840 ACATGGCTGGGCTGCT---TTCACACTTACCCCTGCC-----AAC 1881
 QY 45 ProProProProGluThrSerAsnProAsnProGlySerGlnThrAsnGlnLeuGln 64
 DB 1882 CCACACACCCCGAGGTGTCCAAATCCCAAAAGCCAGAGACGATTACCAACCGCTGCA 1941
 QY 65 TyrLeuLeuArgValIleLeuLeuSerThrLeuThrPheLeuSHISGlnPheAlaTrpProPheGln 84
 DB 1942 TACCTTACCAAGGATGATGATGAGGCTGTGGAACATCGATTGCGCATTCGCG 2001
 QY 85 GlnProValAspAlaValIleLeuLeuAsnLeuProAspTyrTyrIleLeuLeuThrPro 104
 DB 2002 CAGCTGTGATGCTGCTCAAACTGGGCTACCGGATTATCAAAATATATAAGGAGCT 2061
 QY 105 MetAspMetGlyThrIleLeuLeuArgLeuGluAsnAsnTyrTyrTrpAsnAlaGlnGlu 124
 DB 2062 ATGACATGTGGTACTATTAAAGAGAGACTTCAAAACATTTATTATGGCTGCTTCAAG 2121
 QY 125 CysIleGlnAspPheAsnThrMetPheThrAsnCysTyrIleTyrAsnLeuProGlyAsp 144
 DB 2122 TGTATGCAAGATTTAATACATGTTTCACTGTTACCACTTATACACAGGCCACTGAT 2181
 QY 145 AspIleValLeuMetAlaGluAlaLeuGluLeuLeuPheLeuGlnIleLeuAsnGluLeu 164
 DB 2182 GATTTGTGCTTAATAGGCAACAAAGCTGGAAGAAAGATTTCTTACAGAGAGTTGCATCATG 2241
 QY 165 ProThrGluGluThrGluIleMetIleValGlnAlaGlyGly-----ArgGly-- 180
 DB 2242 CCACAGAGAAACAAAGAGCTGTAGTACCATCCCTTAAGAACAGCACAGAGAGGGGCC 2301
 QY 181 -----ArgGlyArgGlySerGluThrGluThrAlaGlyProGlyValSerThr 195
 DB 2302 AAGTTGGCAGCGCTCCAGGGCACTTTACAGTCCCATCAGAGTCCCTGCGCTCTTCT 2361
 QY 196 ValProAsnThrThrGlnAlaSerThrProProGlnThrGlnThrProGlnProAsnPro 215
 DB 2362 GTGTCACACACAGCCCTTATACCTCCACCTGAGATA-----CCT 2403
 QY 216 ProProValGlnAlaThrProHisProPheProAlaValThrProAspLeuIleValGln 235
 DB 2404 ACCACTGTCTCAACATTCCTCCACCCCA-----TCAGTCAATTTCC 2442
 QY 236 ThrProValMetThrValValProProGlnProLeuGlnThrProProValProPro 255
 DB 2443 TCTCCACTTCTCAAGTCTTGAC-----TCTGCTGACCCCGCTCTTGGCT 2490
 QY 256 GlnProGlnProProAlaProAlaProGlnProValGlnSerHisProProIleIle 275
 DB 2491 GTTACTGACGCTCTCCAGCC----- 2511
 QY 276 AlaAlaThrProGlnProValIleThrIleGlyValIleValGlyAlaAspThrThr 295
 DB 2512 -----CAGCCCTTGCCCAAGAAAGCGGTAAGCGGAAAGCAGATCTAC 2559
 QY 296 ThrProThrThrIle-----AspProIleHisGluProProSerIleu 309
 DB 2560 ACCCTTACCTTACAGCATTGTGGCTCTGTTCTCCAGCTTACCCCTCTGAGGAGTCTT 2619

QY 310 ProProGluProPolysThrThrIleLeu--GlyGlnArgArgGluSerSerArgProVal 328
 DB 2620 -----GAGCTTAAGGACAGACCGCTTCCCTTATGATGAGAGAGTGTGGCCATC 2673
 QY 329 LysProProLysAspValProAspSerGlnGlnHisProAlaProGluLysSerSer 348
 DB 2674 AAGCCCCAGCAAGACTTGGCTGACTTCCAGCAACACACAGAGCTCTTAAGAAAGCA 2733
 QY 349 LysValSerGluGlnLeuLysCysCysSerGlyIleLeuLysGluMetPheAlaLysLys 368
 DB 2734 AAGCTTCAGAACAGTTAAACATTGCAATGCAATTTGAAGAGATTACTTAAAG 2793
 QY 369 HisAlaAlaTyrAlaTrpProPheTyrIleProValAspValGluAlaGluLysHis 388
 DB 2794 CATGCTCCATATGCTTGGCTTCTTATTAACCAAGTGATCTTCTGCACTTGGCTTCAT 2853
 QY 389 AspTyrCysAspIleIleLeuHisProMetAspMetSerThrIleLysSerLysLeuGlu 408
 DB 2854 GACTTACATGACATCATTAAGACCCCATGACCTTCAGCACTGTCAAGCGGAGATGGAG 2913
 QY 409 AlaArgLysTyrArgAspAlaGlnGluPheGlyValAspValArgLeuMetPheSerAsn 428
 DB 2914 AACCTGATTACCGGATGACAGAGACTTGGCTGATGATGAGGCTTATGTTCTCCAC 2973
 QY 429 CysTyrLysTyrAsnProProAspHisGluValAlaMetAlaArgLysLeuGlnAsp 448
 DB 2974 TGCTATTAAGTACATACCCCAAGATCAGAGATGCTGTGGCAATGGCACAAAGCTTACAGAT 3033
 QY 449 ValPheGluMetArgPheAlaLysMetProAspGluProGluGluProValIleAlaVal 468
 DB 3034 GTATTTAAGTTCGTTATGCGAAGATCCAGATGACACACTAGAACCAAGGAGCCTTTACA 3093
 QY 469 SerSerProAlaValProProPro---ThrLysValValAlaProProSerSerSerAsp 487
 DB 3094 GTCTTACTGACATGCCCTCGCTGGCCCAATCGCTTACAGAGTCTCCAGAGAGAA 3153
 QY 488 SerSerSerAspSerSerSer-----Asp 495
 DB 3154 AGTAGCAGTGAAGCTCTCTCTGAGAGAGAGAGAGAGATGAGAGAGAGAGAGAA 3213
 QY 496 SerAspSerSerThrAspAspSerGluGluGluArgAlaGlnArgLeuAlaGluLeuGln 515
 DB 3214 GAAAGAGTGAAGCTCAGACTCAGAGAGAGAGAGAGGCTCATCTTAAAGAACTTACAG 3273
 QY 516 GlnGlnLeuLysAlaValIleGluGlnLeuAlaLeuSerGlnProGlnGlnAsnLys 535
 DB 3274 GAAACGCTTGGGCGAGTACATGAACACTGCTGCTGCCAGGGTCCAAATATCCAAAG 3333
 QY 536 ProLysLysLysGluLysAspLysLysGluLysLys-----GluLysHis 551
 DB 3334 CCCAAGAGGAAA---AGAGAGAAAAAGAGAAAAAGAGAAACGAGAGAGAGAACAT 3390
 QY 552 LysArgLysGluGluValIleGluGluAsnLysLysSer---LysAlaLysGluProPro-- 569
 DB 3391 CGAGCGGAGCTGGGGCGGATGAAGAGACAGAGGCGCTTGGGACCCCGCCCACTCAA 3450
 QY 570 ProLysLysThrLysLysAsnAsnSerSerAsnSerValSerLysLysGluProAla 589
 DB 3451 CTTAAGAGTCCAAAGAAAGCAAGTGCAGTGGGCTGCGCTTGAAGGCTTCT 3510
 QY 590 -----PrometLysSerLys 594
 DB 3511 GCCTTGGACCTTCTGAGAGAGATGCGACCAAGCTCCCAAAAAGGCACAAAGACAGCC 3570
 QY 595 ProPro-----ProThr---TyrGluSerGluGluLysLysCysLysProMetSer 611
 DB 3571 CCACCTCCCTGCTTACAGGTTATGATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 3630
 QY 612 TyrGluGluLysArgGlnLeuSerLysAspIleAsnLysLeuProGluGluLysLeuGly 631
 DB 3631 TACGATGAAGCGGACCTGAGGCTTGAGATCAACAATTAATCTGGGAGAGAGCTGGC 3690
 QY 632 ArgValValHisIleIleGlnSerArgGluProSerLeuLysAsnSerAsnProAspGlu 651


```
|||||
Db 3691 CGAGTGTGATTAATCCAGGAGGAGCCCTTTACGTATTCAAACCCAGAGAG 3750
QY 652 TleGlIleAspPheGlnThrLeuLysProSerThrLeuArgGlnLeuGlyAlaLeuCys 671
Db 3751 ATTGAATGATTGTTTGAACACTCAAGCCATCCACTTAGAGAGCTTGA-GGGCTAATG 3809
QY 672 HleLeuLeu 674
Db 3810 CCTTCTCG 3818

RESULT 2
US-09-764-864-372
; Sequence 372, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 372
; LENGTH: 746
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (387)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (646)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (665)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (736)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-864-372

Alignment Scores:
Pred. No.: 1,5e-33 Length: 746
Score: 688.00 Matches: 144
Percent Similarity: 66.40% Conservative: 22
Best Local Similarity: 57.60% Mismatches: 61
Query Match: 18.91% Indels: 24
DB: 10 Gaps: 6

US-09-700-590A-22 (1-688) x US-09-764-864-372 (1-746)
QY 51 SerAsnProAsnLysProLysArgGlnThrAsnGlnLeuGlnTyLeuLeuArgValVal 70
Db 3 TCCAAACCCAGCAAGCCGCGCCCAAGACCAACAGCTGCAGTACATGAGAAATGCGTG 62
QY 71 LeuLysThrLeuTrpLysHisGlnPheAlaTrpProPheGlnGlnProValAspAlaVal 90
Db 63 GTCAAGACGCTCGAAGAACACAGCTTCGCTTCACAGCCCGTGGACGCAATC 122
QY 91 LysLeuAsnLeuProAspLysTyTrpLysAlaLeuLysThrProMetAspMetGlyThrIle 110
Db 123 AAATTGAACCTGCGGATTAATCAATAATTAATAAACCAATGATATGGGACTATT 182
QY 111 LysLysArgLeuGlnLysAsnGlnTyTrpAsnAlaGlnGlnLysIleGlnAspPheAsn 130
Db 183 AAGAAGAGACTAGAAATAATATTATTTGAGTGCAGCAAGTATGAGGACTTCAAC 242
QY 131 ThrMetPheThrAsnGlyTrpLysAsnLysProGlnLysAspIleValIleMetAla 150
Db 243 ACCATGTTTACAAATGTTTACATTATTAACAAGCCACAGATGACATATGCTAATGCC 302
```

```
QY 151 GluAlaLeuGlnLysLeuPheLeuGlnLysIleAsnGlnLeuProThrGlnGlnThrGln 170
Db 303 CAAGCTTTAGAGAAATTTTCTACAAAAAGTGCCCATGATGCCCAAGAGAAATTGAA 362
QY 171 IleMetIleValGlnAlaLysGlnArgGlnArgGlnGlnThrGlnThrAlaLys 190
Db 363 ATTATTACCCCTGCTCCAAAGGNCAAAGTCCAGAGCCGGCT-GCGGAGCCCAAGAC 421
QY 191 ProGlnValSerThrValProAsnThrThrGlnAlaSerThrProGlnThrGlnThr 210
Db 422 GCAGTACACAGAGAGTGGCGGCGGTCTCTGCTCC-----CCAGGAC 469
QY 211 Pro---GlnProAsnProProValGlnAlaThrProHisProPheProAlaValThr 229
Db 470 CCTTTCAAGAGCGTGGCCCAACCGCTCCACAGCCGCC----- 508
QY 230 ProAspLeuIleValGlnThrProValMetThrValProProGlnProLeuGlnThr 249
Db 509 -----GTCACTGCTGCCACCCCTGTACCAACCATCACTGCAAAC-----GTCACG 553
QY 250 ProProProValProProGlnProGlnProProProAlaProAlaProGlnProValGln 269
Db 554 TCGGTCCAGTCCCGCCCACTGCGGCCCACTCTCT-----CCTGCCACA 598
QY 270 SerHisProProIleIleAlaAlaThrProGlnProValLysThrLysGlyValLys 289
Db 599 CCCATGCTCCCGGTGCTCCCTCCACGCCA---CCTGTGTCAGAAAGGCGTTGAA 655
QY 290 ArgLysAlaAspThrThrThrProThrThr 299
Db 656 CGGAAAGCANACACCACTCCACGACT 685

RESULT 3
US-09-764-864-785
; Sequence 785, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 785
; LENGTH: 424
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-864-785

Alignment Scores:
Pred. No.: 2,15e-26 Length: 424
Score: 564.00 Matches: 105
Percent Similarity: 85.71% Conservative: 15
Best Local Similarity: 75.00% Mismatches: 19
Query Match: 15.50% Indels: 2
DB: 10 Gaps: 1

US-09-700-590A-22 (1-688) x US-09-764-864-785 (1-424)
QY 51 SerAsnProAsnLysProLysArgGlnThrAsnGlnLeuGlnTyLeuLeuArgValVal 70
Db 3 TCCAAACCCAGCAAGCCGCGCCCAAGACCAACAGCTGCAGTACATGAGAAATGCGTG 62
QY 71 LeuLysThrLeuTrpLysHisGlnPheAlaTrpProPheGlnGlnProValAspAlaVal 90
Db 63 GTCAAGACGCTCGAAGAACACAGCTTCGCTTCACAGCCCGTGGACGCAATC 122
QY 91 LysLeuAsnLeuProAspLysTyTrpLysAlaLeuLysThrProMetAspMetGlyThrIle 110
Db 123 AAATTGAACCTGCGGATTAATCAATAATTAATAAACCAATGATATGGGACTATT 182
```

```

Oy 111 LysIVsYrGLEuGluuAenAsnTYrTYrTriPasnIaGnIuCYsIIEgInAspPheAsn 130
Db 183 AAGAAAGACACTGAAATAATATATATATTTAGAGTCAAGCGAAATGATGACGAGACTTCAAC 242
Oy 131 ThirMetPheThraAsCYsTYrIIEtyrAsnIysProGluIysAspAplIeValIleuMetAla 150
Db 243 ACCATGTTTACAAATGTTTACATTTATTTATTAACAAGCCCAACAGATGACATATGCTAATAGCC 302
Oy 151 GluAlaIleuGluIysIleuPheIleuGlnIlyIleasnIleuProThrIrgIuGluIuThrGln 170
Db 303 CAACGTTTAAAGAAAATTTTCTCAAAAATACTGGCCAGATGCCCCCAAGGAGGAAGTTGAA 362
Oy 171 IleMetIleValIleGlnAlaIysGlyIAsrGlyIAsrGlyIAsrGlyIuThrGlyIuThrAlaIys 190
Db 363 ATT--ATTACCCTCTGCTCCAA-AGGGGCAAAAGCTGCGAAGCCGGCTGCGGAGAGCCAG 418

RESULT 4
US-09-879-536-516
Sequence 516, Application US/09879536
Patent No. US20020144298A1
GENERAL INFORMATION:
APPLICANT: Endeque, Wilson O.
APPLICANT: Steinmann, Kathleen E.
APPLICANT: Astle, Jon H.
APPLICANT: Burgess, Christopher C.
APPLICANT: Bushnell, Steven E.
APPLICANT: Carroll III, Eddie
APPLICANT: Catino, Theodore J.
APPLICANT: Dertli, Adnan
APPLICANT: Ford, Donna M.
APPLICANT: Lewis, Marcia E.
APPLICANT: Monahan, John E.
APPLICANT: Schlegel, Robert
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
FILE REFERENCE: CCD-257 (US)
CURRENT APPLICATION NUMBER: US/09/879, 536
CURRENT FILING DATE: 2001-09-21
PRIOR APPLICATION NUMBER: US 60/088, 801
PRIOR FILING DATE: 1998-06-10
NUMBER OF SEQ ID NOS: 850
SOFTWARE: PasteSeq for Windows Version 3.0
SEQ ID NO 516
LENGTH: 450
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)-(450)
OTHER INFORMATION: n = A,T,C or G
US-09-879-536-516

Alignment Scores:
Pred. No.: 1,28e-21 Length: 450
Score: 484.50 Matches: 96
Percent Similarity: 75.00% Conservative: 12
Best Local Similarity: 66.67% Mismatches: 27
Query Match: 13.31% Indels: 9
Gaps: 3
DB: 10

US-09-700-590A-22 (1-688) x US-09-879-536-516 (1-450)
Oy 286 LysGIyValIysArgIysAlaAspThrThrThrProThrThrIle----- 300
Db 3 AAAGCGCTAAAGCGGAAAGCAGATCTACACCCCTACACTACAGCAGCATTTGGCTCT 62
Oy 301 --AspProIleHisGluProProSerIeuProProGluIuProIySthThrIysIeu-- 318
Db 63 GGTCTCCAGCTAGCCCTCTCGGAGTCTT-----GAGCTTAAGCAGACGCGCTTCCC 116
Oy 319 GlyGlnAArgArgGluSerSerArgProValIysProProIySlyAspValProAspSer 338
Db 117 CCTATGCGGTAGAGAGGTGTGCGCCCATCAAGCCCAAGGAAAGCTTGCGTGACTCT 176

```

```

Oy 339 GlnGlnHisProAlaProGluLysSerSerLysValSerGluGlnLeuLysCysCysSer 356
Db 177 CAGACACACACCCAGACCTCTTAAGAAAGAAAGAGCTTTCACACACGTTAAACATTGCCAAT 236
Oy 359 GlyIleLeuLysGlnMetPheAlaLysLysHisAlaAlaIleValaIleTrpProPheTyrLys 378
Db 237 GGCATTTTGAAGAGATTACTCTCAACAAACAGTCTGCTCATCTGGCTTTCTTAATAA 296
Oy 379 ProValAspValGluAlaLeuGlyLeuHisAspTyrCysAspIleIleLysHisPromet 398
Db 297 CCAGTGGATGCTTCTGCACCTTGGCTGTCACAGCTACCATACATCACTTAAGACACCCATG 356
Oy 399 AspMetSerThrIleLysSerLysLeuGluAlaArgGluTyrTrpAspAlaGlnGluPhe 418
Db 357 GACCTCAGCAGCTGTCAAGCGGAGATGAGAACCGTGATTACCGGAGATGACAGAGTTT 416
Oy 419 GlyAlaAspVal 422
Db 417 GCTGCTGATGTA 428

RESULT 5
US-09-876-889-243
; Sequence 243, Application US/09876889
; Patent No. US20020076715A1
; GENERAL INFORMATION:
; APPLICANT: Benson, Darin R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR OVARIAN
; TITILE OF INVENTION: CANCER THERAPY AND DIAGNOSIS
; FILE REFERENCE: 210121.466C3
; CURRENT APPLICATION NUMBER: US/09/876,889
; CURRENT FILING DATE: 2001-06-06
; NUMBER OF SEQ ID NOS: 353
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 243
; LENGTH: 429
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-876-889-243

Alignment Scores:
Pred. No.: 2,366-14 Length: 429
Score: 362.50 Matches: 79
Percent Similarity: 68.53% Conservative: 19
Best Local Similarity: 55.24% Mismatches: 24
Query Match: 9.96% Indels: 21
DB: 10 Gaps: 5

US-09-700-590A-22 (1-688) x US-09-876-889-243 (1-429)
Oy 538 LysLysGluLysAspLysLysGluLysLysGluLysGluLysHisLysAspGlyGluGluVal 557
Db 2 AAAAAAGAAAAA--AACAGAAAAAGAGGACAGAGAAACATCTGGCCGCAATTGGGATC 58
Oy 558 GluGluAsnLysLysSer--LysAlaLysGluProPro--ProLysLysThrLysLys 575
Db 59 GATGAAGATGATTAAGGGGCGCTAGGAGCACTGCGCCACTCAGCCCAAGAAATCTAAGAAA 118
Oy 576 AsnAsnSerSerAsnSerAsnVal----- 583
Db 119 GCAGGTGTGTGGGGGTAGCAATGCTACTACACTCAGCCATCTGTGGCTTTGGGACTTCGGA 178
Oy 584 -----SerLysLysGluProAlaPrometLysSerLysProProProthr----- 598
Db 179 GGAAGTACCAACAAGCTAACTAAAGTCTCAAAAAGACAGCTCCACTGTCCTTCCACT 238
Oy 599 ---TyrGlnSerGluGluGluAspLysCysLysPrometSerTyrGlnGluLysArgGln 617
Db 239 GGCTATGATCTTGAAGAGAGAGAGAAAGAAAGACGCGCCATGACTTATGATGAGAGAGCAG 299

```

QY 618 LeuSerLeuasp11leasnlyLeuProGlyGlyLeuGlyArgValValHisIleIle 637
DB 299 TTAAGCTGATATCATAGTAAGTACCTGGGGAAGCTGGGTGAGTAGATATCATC 358
QY 638 GluSerArgGluProSerLeuLysAsnSerAsnProAspGluIleGluIleAspPheGlu 657
DB 359 CAAAGCCAGGGAACCTCTCTACGATTCAATCATCAGAGAAATTGATTGATTGAA 418
QY 658 ThrLeuLys 660
DB 419 ACACTCAAG 427
RESULT 6
US-10-109-886-7
Sequence 7, Application US/10109886
Patent No. US20020119499A1
GENERAL INFORMATION:
APPLICANT: TANABE SEIYAKU CO. LTD.
APPLICANT: TANIGUCHI, Tomoyasu
APPLICANT: MIZUKAMI, Junko
TITLE OF INVENTION: METHOD FOR IDENTIFYING OR SCREENING AGONIST AND
TITLE OF INVENTION: ANTAGONIST TO PPAR
FILE REFERENCE: TANIGUCHI-6
CURRENT APPLICATION NUMBER: US/10/109,886
CURRENT FILING DATE: 2002-04-01
PRIOR APPLICATION NUMBER: 09/514,247
PRIOR FILING DATE: 2000-02-28
PRIOR APPLICATION NUMBER: PCT/JP98/03734
PRIOR FILING DATE: 1998-08-24
PRIOR APPLICATION NUMBER: JP231084/1997
PRIOR FILING DATE: 1997-08-27
NUMBER OF SEQ ID NOS: 10
SOFTWARE: PatentIn version 3.0
SEQ ID NO 7
LENGTH: 7326
TYPE: DNA
ORGANISM: mouse
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(7326)
OTHER INFORMATION: n at position 1131 is unknown.
US-10-109-886-7
Alignment Scores:
Pred. No.: 9.04e-11 Length: 7326
Score: 323.50 Matches: 135
Percent Similarity: 35.88% Conservative: 72
Best Local Similarity: 23.40% Mismatches: 177
Query Match: 8.89% Indels: 193
DB: 12 Gaps: 27
US-09-700-590a-22 (1-688) x US-10-109-886-7 (1-7326)
QY 9 ThrArgLeuArgAsnLeuProValMetGlyAspGlyLeuGluThrSerGlnMetSerThr 28
DB 2008 ACAAGCTTACATAGCAAGGATCTCGGTAAAC----- 2040
QY 29 ThrGlnAlaGlnAlaGlnProGlnProAlaAsnAlaIleSerThrAsnProProPro 47
DB 2041 -----CAGCCAGCTTTTACACAGCTTCTGGGGCTCAGCCCTGTG 2079
QY 48 ---ProGluThrSerAsnProAsnLysProLysArgGlnThrAsnGlnLeuGlnTyrLeu 66
DB 2080 ATTCCACACAGCCAGCTGTGAAGACTCCAAAT----- 2112
QY 67 LeuArgValValLeuLysThrLeuTyrHisGlnPheAlaTyrProPheGlnGlnPro 86
DB 2113 -----GGGCCCCGTGCTTGTCCA 2130
QY 87 ValAspAlaValLysLeuAsn-----LeuProAspTyrTyrLysIleIleLysThrPro 104
DB 2131 GTGAATGATGATGAGTTTCTCAAGAGATGATGATTCATTT-----AACCCA 2175

QY 105 MetAspMetGlyThrIleLys-----LysArg 113
DB 2176 ATGTCTCCAGGAAACCTTCCAGTCCACAGGACCCATGGACCTGTGACGCTCCCT 2235
QY 114 LeuGluAsnAsnTyrTyrTrpAsnAlaGlnGluCysIle----- 126
DB 2236 ATGAACCATCTCTGCAAGATGAACAGATGCGCTCAGTTCGGGTATGGCATTTCTCT 2295
QY 127 -----GlnAspPheAsnThrMetPheThrAsnCysTyrIleTyrAsnLysPro 142
DB 2296 TCAGGATGCTCAGGCTCCAAATATGATGGGACATCAGCC-----AACAC-- 243
QY 143 GlyAspAsp11leValLeuMetAlaGlnIleValGlnAlaLysGlyArgGlyArg 161
DB 2344 -----ATTATGGCCAGGACCTACTACAGAACGATTTTGGCCACAG-- 2385
QY 162 AsnGluLeuProThrGluGluThrGluIleMetIleValGlnAlaLysGlyArgGlyArg 181
DB 2386 AACAGATTTCATCA--TCCAGTGGGGCAATGATGTGAACAGTGTGGCATGGGCA 2442
QY 182 GlyArgLysGluThrGlyThrAla-----LysProGlyValSer----- 194
DB 2443 CCAAGACCCAGGACAGGTGTTTACAGAGGTCAAGAACTGGAGTGTCTTCCCTAACCT 2502
QY 195 -----ThrValProAsnThrThrGlnAlaSerThrProGlnThrGlnThrPro 211
DB 2503 CTGAACATGCTGAGCCCAAGGACCCAGCTGCTTGGCCACAGTGAACAGTCAACCA 2562
QY 212 ---GlnProAsnProProPro-----ValGlnAlaThrProHisPro 224
DB 2563 TTGCACCCAGCTCAGCTCTCTGTTCCAGAGCTGTGAGTGGCTCTCTCCAAACATCA 2622
QY 225 ---PheProAlaValThrProAspLeuIleValGlnThrProValMetThrValValPro 243
DB 2623 ACGGCACACAGAAATGACCT-----CTCAGCCAGCAGCTCCCACTACGCCA 2670
QY 244 ProGlnProLeu-----GlnThrProProProValPro----- 254
DB 2671 TCTACTCTGTGTCTGTGAGGACAGTCTTCAACCACTCTGTGATGCCAGCT 2730
QY 255 -----ProGln 256
DB 2731 GCCCAACACAGATACCCCTACAGTCCAGGACAGCAGCAGCAGCTGAGTACTCCACG 2790
QY 257 ProGlnProProProAlaPro-----AlaPro 285
DB 2791 CCTCAGACCCCACTGACGACCATCTGTGGCTACTCTCAGTCACTACAGCAGCAACCA 2850
QY 266 GlnProValGlnSerHisProProIleIle----- 275
DB 2851 ACGCTGTGATATCTAGCCACTGCGCACACCGCTTCTCAGGACAGCAGCATTTGAT 2910
QY 276 -----AlaAlaThrProGlnProValLysThrLysLysGlyValLysArgLysAlaAsp 293
DB 2911 AATAGATGCTCTACTCTCTCACTGACAGAGTGTGAACCAAGTTCCACAGCAGCAGGA 2970
QY 294 ThrThrThrPro-----ThrThrLeuAspProIleHisGlu 305
DB 2971 CCCGATGTGCTGCTGGAATGAAGACAGAGCTGACAGCATATGCTGAGCTGAA 3030
QY 306 ProProSerLeuProProGluProLysThrThrLysGlnArgArgGlnSerSer 325
DB 3031 CTTACTGATTCGAAGGGGGAACCTCGTGTGAGATGATGAAGAGAGATTTCAGAGTTCT 3090
QY 326 ArgProValLys-----ProProLysLysAspValProAspSerGlnGln 340
DB 3091 TCCCAAGTAAAGAAAGACAGATACGACAGACAGAAAGTCAAGCAATGGAATGAA 3150
QY 341 HisProAlaProGlu-----LysSerSerLys 349
DB 3151 GAAAGAAACCTGAATGAATAAGTGAAGCTAAAGAGAAAGAAAGAACTGTGAAGCAGC 3210
QY 350 ValSerGluGlnLeuLysCysCysSerGlyIleLeuLysGluMetPheAlaLysLys--- 368

```
Db 3211 ACACCTCACATCAATCATCTCTCCAGCCAGCCAAAATACTTTAAACCCAGAG 3270
OY 369 -----HisAlaIaTYr-----AlaItrp 374
Db 3271 CTACCCAGCAGCATTATCCCACTTGAAGACCTTTCACAGAGAGCCAGAGCTTTG 3330
OY 375 ProPheTYrIysProValAspValGluAlaLeuGlyLeuHISAspTYrCysAspIleIle 394
Db 3331 CTTTGGTCAAGCTGTGATCTCAGCTCAGGAATCCAGATTAATTTGATATAGTGTG 3390
OY 395 LysHisProMetAspMetSerThrIleIysSerIysLeuGluAlaArgGluTYrArgAsp 414
Db 3391 AAGATCTTATGAGACCTTTCTTCAACATCAAAAGCAGACAGGCGCAATATCAAGAA 3450
OY 415 AlaGluGluPheGluAlaAspValArgLeuMetPheSerAsnCysTYrIysAspPro 434
Db 3451 CCCTGGCAGTATGTCAGTATGCTGAGCTTATGTTCAACATTCCTGCTATATATGCT 3510
OY 435 ProAspHisGluValAlaValAlaMetAlaArgIysLeuGluAspValPheGlu 451
Db 3511 AAAACGTCCCGTGTATATTAATTTTGCAGTAAACTTGCAGAGAGCTTTGAA 3561

RESULT 7
US-09-815-343-213/c
; Sequence 213, Application US/09815343
; Patent No. US2001005596A1
; GENERAL INFORMATION:
; APPLICANT: Meagher, Madeleine
; APPLICANT: Xu, Jiangchun
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; FILE REFERENCE: 210121.504
; CURRENT APPLICATION NUMBER: US/09/815.343
; CURRENT FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 213
; LENGTH: 547
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1) ... (547)
; OTHER INFORMATION: n = A,T,C or G
US-09-815-343-213

Alignment Scores:
Pred. No.: 5,08e-11 Length: 547
Score: 308.50 Matches: 75
Percent Similarity: 60.48% Conservative: 26
Best Local Similarity: 44.91% Mismatches: 35
Query Match: 8.48% Indels: 31
Gaps: 8
DB: 10

US-09-700-590A-22 (1-688) x US-09-815-343-213 (1-547)
OY 491 AspSerSerSerAspSerAspSerSerThrAspAspSerGluGluGluArgAlaGlnArg 510
Db 516 GAGGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 457
OY 511 LeuAlaGluLeuGluGluGluLeuIysAlaValHISGluGluLeuAlaIleuSerGln 530
Db 456 TTAGCAGAACTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 397
OY 531 ProGluGluAsnIysProIysIysIysGluIysAspIysIysGluIysIysIys----- 548
Db 396 GGTCAATATCCAAAGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 340
OY 549 -----GluIysHisIysLeuArgIysGluGluValGluGluAsnIysIysSer---LysAla 565
Db 339 AAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 280
```

```
OY 566 LysGluProPro---ProIysIysThrIysIysAsnAsnSerSerAsnValSer 584
Db 279 CCCCAGCCACTCACTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 220
OY 585 LysIysGluProAla----- 589
Db 219 GCTTAGGCCCCCTTCTGCTTTGAGACCTTTCAGAGAAAGTGGACCAAGCTCCCCAAAAG 160
OY 590 ProMetIysSerIysProPro-----ProThr---TYrGluSerGluGluIysAspIys 606
Db 159 GCCAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 100
OY 607 CysIysProMetSerTYrGluGluIysArgGluIleuSerIleuAsnIysLeuPro 626
Db 99 AGCAGAGCCCATGAGTTACGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 58
OY 627 GluIysIysLeuGluIysArgVal 633
Db 57 GGGGGGGCCGCAAGGCGAAT 37

RESULT 8
US-10-109-886-9
; Sequence 9, Application US/10109886
; Patent No. US2002011949A1
; GENERAL INFORMATION:
; APPLICANT: TANABE SEIYAKU CO. LTD.
; APPLICANT: TANIGUCHI, Tomoyasu
; APPLICANT: MIZUKAMI, Junko
; TITLE OF INVENTION: METHOD FOR IDENTIFYING OR SCREENING AGONIST AND
; FILE REFERENCE: TANIGUCHI=6
; CURRENT APPLICATION NUMBER: US/10/109.886
; CURRENT FILING DATE: 2002-04-01
; PRIOR APPLICATION NUMBER: 09/514,247
; PRIOR FILING DATE: 2000-02-28
; PRIOR APPLICATION NUMBER: PCT/JP98/03734
; PRIOR FILING DATE: 1998-08-24
; PRIOR APPLICATION NUMBER: JP231084/1997
; PRIOR FILING DATE: 1997-08-27
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9
; LENGTH: 8147
; TYPE: DNA
; ORGANISM: human
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (819) .. (8147)
US-10-109-886-9

Alignment Scores:
Pred. No.: 8.49e-10 Length: 8147
Score: 308.00 Matches: 109
Percent Similarity: 37.80% Conservative: 49
Best Local Similarity: 26.08% Mismatches: 132
Query Match: 8.46% Indels: 128
Gaps: 19
DB: 12

US-09-700-590A-22 (1-688) x US-10-109-886-9 (1-8147)
OY 148 LeuMetAlaGluAla---LeuGluIysLeuPheLeuGluIysIleAsnGluLeuProThr 166
Db 3165 ATGATGGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3221
OY 167 GluGluThrGluIleMetIleValGluAlaIysGluIysArgIysGluIysArgIysGluThr 186
Db 3222 TCCAGCGGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3269
OY 187 GlyThrAlaIys-----ProGlyValSerThrVal----- 196
Db 3270 GAGCGTGCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3329
```

QY 197 ProAsnThrThrlAlaSerThrProProGlnThrlThrPro---GlnProAsnPro 215
Db 3330 CCTCAGGCCGCGCTACCTGCTCCGAGTGCACAGTCCAGCTGCACCAACGCG 3389
QY 216 ProProValGlnAlaThrProHisProPheProAlaValThrProAspLeuIleValGln 235
Db 3390 CCTCTGCTCCACGGCT-----GCTGGCATGCTCATCTCCAGCACACG 3434
QY 236 ThrProValMetThrValValProProGlnPro---LeuGlnThrProProValPro 254
Db 3435 ACACCA---CCTGGATGATGATCTCTCCACGACGACACTCCCACTACCACTCCT 3491
QY 255 -----ProGlnProGlnProProProAlaProAlaProGlnPro 267
Db 3492 GTGTGCTTCCGGGAGAGCTCCACCCCACTGCTGCTCAGTGCCTGCTGCTCA 3551
QY 268 ValGlnSerHisProProIleIleAlaAla-----ThrProGln----- 280
Db 3552 ACCCAGAGCAGCCCTCAGTCCAGGCGAGCGAGCCAGGCCAGGTGACCCGAGCTCAA 3611
QY 281 ---ProValThrLysLysGlyValLysGlyAlaAspThrThrThrProThrThr 299
Db 3612 ACCCCAGTACGCCCCGCTGAGCTACCCCTCAGTATCGACGACAGCCGAGC--- 3668
QY 300 IleAspProIleHis---GlnProProSerLeuProProGlnProLysThrLysLeu 318
Db 3669 ---CCTGTGACGCGCCGCTCTGAGCAGACCGCTTCCAGCAGCAGCAGCAGT 3722
QY 319 GlyGlnArgArgLysSerSerArgProVal-----LysProPro 331
Db 3723 GATACAGAGTCCCTACCCCTCTCGGTGCGCAGCGAGAAACCAATCCCGACAGCCA 3782
QY 332 LysLysAspValPro-----AspSerGlnGln 340
Db 3783 GAGCTGACCTACTCTGCTGCGAAATGAATGAAGAGAGAGAGAGAGAGAGAGAGAG 3842
QY 341 HisProAlaProGlnLys-----SerSerLysValSerGlnLysLysCys 356
Db 3843 GATCTGTGTAATCCAAAGGGAGGCCAGCGTCTGATGATGAGAGAGATTTCAGGA 3902
QY 357 CysSerGlyLysLeuLysGlu----- 363
Db 3903 GCTTCCCAAGTTAAAGAAAGACAGATAGCAGAGCAGAAATCAGAACATGGAAGTG 3962
QY 363 ----- 363
Db 3963 GATGAAAAAAGAACTGTAAGTGAAGTAAAGTAAAGAGAGAGAGAGAGAGTAAAC 4022
QY 363 ----- 363
Db 4023 GGCAGAGCTCTGATGACATCTCTTCGAGCGCGCAAAAAATCTTTAAACAGAG 4082
QY 364 -----MetPheAlaLysLysHisAlaAlaTyr-----Ala 373
Db 4083 GAGTTAGCCAGGCCCTCAGTACCCAAACCTAGAGCACTGATTCACAGAGAGAGAGTCA 4142
QY 374 TyrProPheTyrLysProValAspValGlnAlaLeuGlyLeuHisAspTyrCysAspIle 393
Db 4143 TTAACCTTTCGCGAGCTGATGATCCAGCTCTCGAATCTCCAGATTTTGAATC 4202
QY 394 IleLysHisProMetAspMetSerThrIleLysSerLysLeuGlnAlaArgGlnTyrArg 413
Db 4203 GTAAGATATCCATGAGCACTCTCCACATCAAGCGAAGCTGGACAGAGGCAATACAA 4262
QY 414 AspAlaGlnGlnPheGlyAlaAspValArgLeuMetPheSerAsnCysTyrLysTyrAsn 433
Db 4263 GAGCCCTGAGATGAGAGAGAGAGAGAGCTGCTGCTCAATGCTGCTGCTTATATAT 4322
QY 434 ProProAspHisGlnValAlaAlaMetAlaArgLysLeuGlnAspValPheGln 451
Db 4323 CGCAAGACATCCGAGTCTATAGTTTGCAGTAAAGCTTGCAGAGGCTTTGAG 4376
RESULT 9

US-09-783-590-1966
; Sequence 1966, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
; APPLICANT: Dillion, Patrick J.
; APPLICANT: Haseltine, William A.
; APPLICANT: Li, Haodong
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
; FILE REFERENCE: PO-16.2C1
; CURRENT APPLICATION NUMBER: US/09/783,590
; PRIOR FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: 08/420,856
; PRIOR FILING DATE: 1995-04-12
; PRIOR APPLICATION NUMBER: 08/346,731
; NUMBER OF SEQ ID NOS: 12485
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1966
; LENGTH: 358
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; LOCATION: (10)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (41)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (102)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (121)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (158)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (180)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (186)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (203)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (225)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (237)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (329)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (331)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (334)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (336)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (339)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature


```

FEATURE:
; OTHER INFORMATION: MAP TO AC005529.7
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 0.94
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 5.6
US-09-864-761-2927

Alignment Scores:
Pred. No.: 1.04e-07 Length: 1958
Score: 262.50 Matches: 126
Percent Similarity: 38.61% Conservative: 96
Best Local Similarity: 21.91% Mismatches: 268
Query Match: 7.21% Indels: 85
DB: Gaps: 20

US-09-700-590a-22 (1-688) x US-09-864-761-2927 (1-1958)
QY 138 ILETyrAsnLysProGlyAspAspIleValIleuMetAlaGluAlaLeuGluLysLeuPhe 157
DB 272 GTGTGCGAAGAGTCTGAGAAGAACTGTGATTTGAGAGAAACAGACAGAGAGG----- 325
QY 158 LeuGluLysIleAsnGluLeuProThrGluGluThrGluIleMetIleValGluAlaLys 177
DB 326 --ACCCAGTGAAGTGAAGTGAAGTGAAGGAGAAAGAGCCAAAGAGAGAGAG 382
QY 178 GLyArGlyAspGlyAspGlyAspGlyAspGlyAspGlyAspGlyAspGlyAspGlyAsp 197
DB 383 GGCAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 427
QY 198 AsnThrThrGlnAlaSerThrProProGlnThrGlnThrProGlnPro----- 213
DB 428 ---GAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 484
QY 214 AsnProProProValGluAlaThrProHisProPheProAlaValThrProAspLeuIle 233
DB 485 AAGTCAACAGTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 544
QY 234 ValGlnThrProValMetThrValValProProGlnProLeuGlnThrProProProVal 253
DB 545 GAAGCAAAATCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 604
QY 254 ProProGlnProGlnProProProAlaProAlaProAlaProAlaProAlaProAlaPro 273
DB 605 GAGGCAAG---TCACCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 661
QY 274 IleIleAlaIleAlaThrProGlnProValLysThr-----LysLysGlyValLysArgLys 291
DB 662 GCTGAGTCAAGTCTCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 721
QY 292 AlaAspThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 310
DB 722 GCTGAGGCGCAAGTCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 781
QY 311 ProGluProLysThrThrLysLeuGlyIleArgArgLysSerSerArgProValLysPro 330
DB 782 GCTGAGGCGCAAGTCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 841
QY 331 ProLysLysAspValProAspSerGlnGlnHisProAlaProGluLysSerSerLysVal 350
DB 842 GAAAAGGCGCAAGTCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 901
QY 351 SerGluGlnLeuLysCysSerSerGlyIleLeuLysGluMetPheAlaLysLysHisAla 370
DB 902 GAG-----AAAGAAAGAGGCC 916
QY 371 AlaTyrAlaTyrProPheTyrLysProValAspValGluAlaLeuGlyLysHisAspTyr 390
DB 917 AAGTCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 958
QY 391 CysAspIleIleLysHisProMetAspMetSerThrIleLysSerLysLeuGluAlaArg 410

```

```

DB 959 -----AAGTCCCT-----GAGAGGCCAAGTCCCGAGAGAGAGAGAGAGAGAGAG 994
QY 411 GLyTyrArgAspAlaGlnGluPheGlyAlaAspValArgLeuMetPheSerAsnCysTyr 430
DB 995 GAGCAAAAGTCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1054
QY 431 LysTyrAsnProProAspHisGluValValAlaMetAlaArgLysLeuGlnAspValPhe 450
DB 1055 AAGGCCAAGTCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1114
QY 451 GluMetArgPheAlaGluMetProAspGluProGluGluProVal----- 465
DB 1115 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1171
QY 466 -----ValAlaValSerSerProAlaValProProProThrLysValAlaProPro 483
DB 1172 TCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1221
QY 484 SerSerSerAspSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 503
DB 1232 GCCAAGACTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1282
QY 504 GluGluGluArgAlaGlnArgLeuAlaGluLeuGlnGluGlnGluLysAlaValHisGlu 523
DB 1283 GAAAAGGCCAAAGCCCTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1342
QY 524 GlnLeuAlaAlaLeuSerGlnProGlnGlnAsnLysProLysLysGlyLysAsp--- 542
DB 1343 CTGAAGGCGAGTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1402
QY 543 -----LysLysGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLys 560
DB 1403 CCACTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1462
QY 561 LysLysSerLysAlaLysGluProProProLysLysThrLysLysAsnAsnSerSerAsn 580
DB 1463 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1510
QY 581 SerAsnValSerLysLysGluProAlaProMetLysSerLysProProProThrTyrGlu 600
DB 1511 -----ACCAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1561
QY 601 SerGluGluGlu-----AspLysCysLysProMetSerTyrGluGluLysArgGln 617
DB 1562 GAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1621
QY 618 LeuSerLeuAspIleAsnLysLeu-----ProGlyGluLysLeuGlyArgValHis 635
DB 1622 GAGGCTGAAGTAAAGAAAAGTCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1681
QY 636 IleIleGlnSerArgGluProSerLysLys-----AsnSerAsnProAspGlu 651
DB 1682 GTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1741
QY 652 IleGluIleAspPheGluThrLysProSerThrLeuArgGluLeuGlyAlaLeuCys 671
DB 1742 GCCAAGGCCAAG---GAAGCCAGCAAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1786
QY 672 HisLeuLeuPheAlaGluGluLysGluThrPheLysLeuArgLys 686
DB 1787 -----GCACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1819

```

```

RESULT 12
US-09-808-880-1/c
; Sequence 1, Application US/09808880
; Publication No. US2003027287A1
; GENERAL INFORMATION:
; APPLICANT: Betlach, Mary C.
; APPLICANT: Shah, Sanjay Krishnakant
; APPLICANT: McDaniel, Robert
; TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE

```

FILE REFERENCE: 30062-20029.00
CURRENT APPLICATION NUMBER: US/09/808,880
CURRENT FILING DATE: 2001-03-14
PRIOR APPLICATION NUMBER: US/09/428,517
PRIOR FILING DATE: 1999-10-28
PRIOR APPLICATION NUMBER: 60/120,254
PRIOR FILING DATE: 1999-02-16
PRIOR APPLICATION NUMBER: 60/106,100
PRIOR FILING DATE: 1998-10-29
NUMBER OF SEQ ID NOS: 12
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO: 1
LENGTH: 50937
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Recombinant DNA
US-09-808-880-1

Alignment Scores:
Pred. No.: 0.000109 Length: 50937
Score: 236.00 Matches: 162
Percent Similarity: 28.41% Conservative: 63
Best Local Similarity: 20.45% Mismatches: 276
Query Match: 6.49% Indels: 291
Gaps: 30

US-09-700-590A-22 (1-688) x US-09-808-880-1 (1-50937)

QY 7 ProGlyThrArgLeuArgAsnLeuProValMetGlyAspGlyLeuGluThrSerGlnMet 26
DB 18063 CCAGCGAGGCTCTTCTGCTGCTCCAGCAGACGACGCGTCTCCACTGTGCGATCG 18004
QY 27 SerThrGlnAlaGlnAla----- 33
DB 18003 CCAGCGCATGAGCGTGCAGGCGTCTCACTCCACGAGTTGGATTGAGAGAACGCTG 17944
QY 34 -----Gln 34
DB 17943 CGATGGCGGTGGCGGTGATGATGCATGCTGCGCGGACGCGGATCCCGGTGG 17884
QY 35 ProGlnProAlaAsnAla-----AlaSerThrAsnProProPro 48
DB 17883 CCCGCGCAGCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 17824
QY 49 GlnThrSerAsnProAsnLysProLysArg-----Gln 59
DB 17823 ACCCGCGCTCAGCGCGGACGCTCTGCGCGGACCATGCCGAGAACCAACGCGCGCTGAG 17764
QY 60 -ThrAsnGlnLeuGlnThrLeuLeuArgValLeuLeuThrLeuThrLysHis----- 77
DB 17763 CAGCGACCAACTCCACCACTCTC-----CTCTGCGCGGACACGAG 17722
QY 78 -GlnPheAlaTrpProPheGlnGlnProValAspAlaValLysLeuAsnLeuProAspTy 97
DB 17721 ACAACCCCGCAACCGCTTCAACCAACCC-----CGACGACTCTCTTCAACCGCA 17671
QY 97 rTyTyLysIleIleLysThrPrometAspMetGlyThrIleLysLysArgLeuGluAsn 117
DB 17670 AGCCCAACCCCTGCGCGCGCGCTGACCTGGCC-----CTCCACAACAG 17626
QY 117 nTyTyTyTrpAsnAlaGlnGluCysIleGlnAspPheAsnThrMetPheThrAsnCy 137
DB 17625 C----- 17625
QY 137 rIleTyAsnLysProGlyAspAspIleValLeuMetAlaGluAlaLeuGluLysLeuPh 157
DB 17625 ----- 17625
QY 157 eleuGlnLysIleAsnGlnLeuProThrGlnGluThrGlnIleMetIleValGlnAla 177
DB 17624 -----ACGCACTCCGCGGAGGTCACTGATCA 17599

QY 177 sGlyArgGlyArgGlyArgLysGluThrGly-ThrAlaLys-ProGlyValSerThrVal 196
DB 17598 ACCGACGGGAGCGGCGAAGAACCGGTGACGAAGATTCCTCCATCATCAGCA 17539
QY 197 ProAsnThrThr-----GlnAlaSerThrProGlnThrGln 209
DB 17538 CCGCGACGAACGCTTCAACACGACCCGATCAGCCATCAGCGCAAGACGCTCG 17479
QY 210 ThrProGlnProAsnProPro----- 216
DB 17478 GATTCATCGCCCGCACCCCGACCGGACACTCCGCGCGGACGACCCACCGACGCA 17419
QY 217 -----ProValGlnAlaThrProHisProPheProAla--ValThrProAsp 231
DB 17418 TCCTCTACACGACCATTCCTCCACGCGACACTCGTGCAGGACGACCAACCGCGAC 17359
QY 232 LeuIleValGlnThrProValMetThrValValProPro----- 244
DB 17358 GACGACCGCCCAACGATCAAGAAACGATTCGCGCGCATACCGCTGCCCCCAG 17299
QY 245 -----GlnProLeuGlnThrProPro----- 251
DB 17298 TGCCCGACACCGCGATTTGAGAGAAACACCAACAGCTCTCAGACACGAGATCCA 17239
QY 252 -----ProValProProGlnProGlnPro--ProProAlaProAlaProGlnPro 267
DB 17238 CCAACTCACCATGATTCACCGGACCGGACCTTGCGCGCATCATCATGCGCTCTCT 17179
QY 268 ValGlnSerHisProProIleIleAlaAlaThrPro--GlnProValThrLysLys 286
DB 17178 GACACAGATCTCCCGCAAGGCGTGAAGTGAAGAACACACCGGATGAAACCGCGG 17119
QY 287 GlyValLysArgLysAlaAspThrThrThrProThrThr--LysProIleHisGlu 305
DB 17118 TCACCGGCTCACCCAGATCCGACACACACGCGGACCGCACAGATACGACATCAC 17059
QY 306 ProProSerLeuProProGlnProLysLysLysLeu----- 318
DB 17058 AGCGCCGACCGACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 16999
QY 319 -----GlnGlnArgArgGlnSerSerArgProValLysPro 330
DB 16998 CACGACGCGCATCGCACGACGACGCGCTCAAGAACCAATGATCCGCCACCA 16939
QY 331 ProLysLysAspVal-----ProAspSerGlnGlnHisProAlaProGluLysSer 347
DB 16938 CCAACCAACCGGCGGATGATCGACCCCAACCAACCAACCAACCGGATGATCGGCTCC 16879
QY 348 SerLysValSerGlnGlnLeuLysCysCysSerGlyIleLeuLysGlnMetPheAlaLys 367
DB 16879 ----- 16879
QY 368 LysHisAlaAlaTrpAlaTrpProPheTyLysProValAspValGluAlaLeuGluLeu 387
DB 16878 ---CAGCGCGCGCCCAACCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 16840
QY 388 HisAspTyCysAspIleIleLysHisPrometAspMetSerThrLysSerLysLeu 407
DB 16839 CAGCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 16798
QY 408 GluAlaArgGluTyArgAspAlaGlnGluPheGlyAlaAspValArgLeuMetPheSer 427
DB 16798 ----- 16798
QY 428 AsnCySTyTyLysTrpAsnProProAspHisGluValAlaAlaMetAlaArgLysLeuGln 447
DB 16797 -----CCACTGATCTCTCA 16783
QY 448 AspValPheGluMetArgPheAlaLysMetProAspGluProGluGluProValValAla 467
DB 16782 -----AACCAAGCGG--ACGCCCAACCAACCGACGACCGCG 16750
QY 468 ValSerSerProAlaValProProProThrLysValValAlaProProSerSerSerAsp 487

```

Db 16749 GGCACACCTCTCATCAACCCCGACGAGTCCACGACCCCACTCGGAT 16690
Qy 488 SerSerSerSerSerSerSerSerSerSerThraSpSerSerGluGluArg 507
Db 16689 GCTCAGAACCCAGAACCCCGAAACCCACACTGGCGCTGCACCGGATCAATCA 16630
Qy 508 AlaGlnArgLeuAlaGluLeuGlnGlnLeuLysAlaValHisGluGlnLeuAla 527
Db 16629 CACCGGCTCGAAGAACCAACCAACCGGCTGCACGACCAACCGCGGCC 16570
Qy 528 LeuSerGlnProGlnGlnAlaSerPro-----LysLysLysGlnLysAsp 542
Db 16569 CCTCACCACCCGCGAGATCAACCAACGCTGCGCAACACGACGCGGCAAC 16510
Qy 543 LysLysGlnLysLysLys-----GlnLysHisLysArgLysGluGluValGlu 558
Db 16509 CGCGGGAGAACAGAAATGCTCCGAGTCCCGCGATCTCCACAGACAAAGAC 16450
Qy 559 GluAlaLysLysSerLysAlaLysGluProProLysLysLysLysAlaSer 578
Db 16449 CCGCGCTCACAACACGCGCGCGAGACCTCGCATTAAGCCGCGGTCAGAGCGG 16390
Qy 579 SerAsnSerAsnValSer--LysLysGluProAlaProMetLysSerLysProPro 597
Db 16389 GATCGAGCTCCAGACACTGACTCGACCCCGCGCTGACGACGACCAACCCG 16330
Qy 598 ThrTyrGluSerGluGluGluAspLysCysLysProMetSerTyrGluGluLysArg 617
Db 16329 CAACCAACCATCATCCGCCAAGCCCGCGGACCAACCAACCAACCAACCA 16273
Qy 618 LeuSerLeuAspLeuAlaSerLysLeuProGluLysLeuGluArgValHisLeu 637
Db 16272 -GCACCTCATCAAGCGCGCGCGGAAAG----- 16246
Qy 638 GlnSerArgLupProSerLeuLysAsnSerAsnProAspGluGluLysAspHeu 657
Db 16245 GCTTCCACTCCGACCGGAT-----AACGCACTCTTAAGCGCTGATCCACCCG 16192
Qy 658 -----ThrLeuLysProSerThr 663
Db 16191 GCCCGCCGACGACGACCAAGCCGCAACA 16162

RESULT 13
US-09-938-842A-1265
; Sequence 1265, Application US/09938842A
; Patent No. US0020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Wang, Xun
; APPLICANT: Kieps, Joel
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIPT300-3
; CURRENT APPLICATION NUMBER: US/09/938, 842A
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227, 866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264, 647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300, 111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 1265
; LENGTH: 1140
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-1265

Alignment Scores: 9.08e-06 Length: 1140
Pred. No.: 226.00 Matches: 86

```

```

Percent Similarity: 35.87% Conservative: 27
Best Local Similarity: 27.30% Mismatches: 117
Query Match: 6.21% Indels: 85
DB: 9 Gaps: 13

US-09-700-590A-22 (1-688) x US-09-938-842A-1265 (1-1140)
Qy 191 ProGluValSerThrVal-----ProAsnThrGlnAlaSerThr-Pr 205
Db 102 CCGTCCACCAACCACTTGTATTTGATTTGTAATCCGCGGACCACTTCGCGGACCA 161
Qy 205 oProGlnThrGlnThrPro-----GlnProAsnProProProValGlnAlaThrPr 222
Db 162 TCCCAACCGGCACTCACCACCAATTCAACGACACCGCGCAACATACCAAC 221
Qy 222 oHisProPheProAlaValThrProAspLeuLeuValGlnThrProValMetThrVal 242
Db 222 GCTTCCACCAACAGACATCAACCACTCCAGTATGACACCAACCA-----CGAGACT 275
Qy 242 LProGlnProLeu-----GlnThrProProProValProPr 255
Db 276 TCTCCAAACCCCTTACCGGACCGGTGTCTCTCCGCAACCAACCACTCTCTCCACC 335
Qy 255 oGlnProGlnProProProAlaProAlaProGlnProValGlnSerHisProProLeu 275
Db 336 AGATTATCAACCAACCACTCCACCACTTATTAACCTTCACTACCTCTCTCCACG 395
Qy 275 eAlaAlaThrProGlnProAlaLysThrLysLysGluValLysArgLysAlaSerThr 295
Db 396 TATTACACCAACCACTCACTA-----GGGAC 422
Qy 295 rThrProThrThrLeaAspProIleHisGluProProSerLeuProProGluProLys 315
Db 423 CACACCAACCGGACTCTCTCCAAACCTTAACCGGCGCTGTCTCT--CCACAAAC 479
Qy 315 rThrLysLeuGluGlnArgArgLysSerArgProValLysProProLysAspVal 335
Db 480 CACG-----CAACCTCTCCACCAACCACTACAC 509
Qy 335 LProAspSerGlnGlnHisProAlaProGluLysSerSerLysValSerGlnLeuLys 355
Db 510 ACCACTATCACT-----CCTCGGTGGATATATGTCGAAGATATACCGAAGCTAA 563
Qy 355 sCysCysSerGluLysLeuLysGluMetPheAlaLysLysHisAlaLysArgLys 375
Db 564 GATATGCGCGAATCTTACCAATTAGTGAACGCTTCTTAACCACTGGAAGAGCAAC 623
Qy 375 oPheTyrLysProValAspValGluAlaLeuGluLysHisAspTyrCysAspIleLeu 395
Db 624 A-----TGTGTTCATCATTCG 641
Qy 395 sHisProMetAspSerThrLys-----SerLysLeuGluAlaArgLys 412
Db 642 AAATGTCTGATTTTATGACAGTACTGTTTGTGAATCCGTAGGACCAACGCTTT 701
Qy 412 rArgAspAlaGlnGluPheGluLysAspValArgLeuMetPheSerAsnCysTyrLys 432
Db 702 TTCCCTCTCCCAATTTCCG-----ATCTCTTAAAGTTGT----- 741
Qy 432 rAsnProProAspHisGluValAlaMetAlaArgLysLeuGlnAspValPheGlu 452
Db 742 -----GGTCGAGAGATTCCACAGAGCTTTAAGCTG 770
Qy 452 rArgPheAlaLysMetProAspGluProGluGluProValAlaLysSerSerProAl 472
Db 771 T-----CCGCGTCACTCAACCAATTCACCAACCA 803
Qy 472 aValProProProThrLysValAlaAlaProProSerSerSer 486
Db 804 ACTTCTTCAACCACTTAAACCGCTCCGCGCAACAAACA 846

RESULT 14
US-09-887-576-652

```

```

; Sequence 652, Application US/09887576
; Patent No. US20020144047A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, P.
; APPLICANT: Brown, D.
; APPLICANT: Chang, H.
; APPLICANT: Zhu, T.
; APPLICANT: Han, B.
; APPLICANT: Wang, X.
; APPLICANT: Cooper, Bret
; TITLE OF INVENTION: Promoters for regulation of plant expression
; FILE REFERENCE: 1360.001US1
; CURRENT APPLICATION NUMBER: US/09/887,576
; CURRENT FILING DATE: 2001-06-25
; PRIOR APPLICATION NUMBER: US 60/213,848
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 60/214,087
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 60/258,692
; NUMBER OF SEQ ID NOS: 875
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 652
; LENGTH: 1140
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; US-09-887-576-652

```

```

Alignment Scores:
Pred. No.: 9,08e-06 Length: 1140
Score: 226.00 Matches: 86
Percent Similarity: 35.87% Conservative: 27
Best Local Similarity: 27.30% Mismatches: 117
Query Match: 6.21% Indels: 85
DB: 10 Gaps: 13
US-09-700-590A-22 (1-688) x US-09-887-576-652 (1-1140)

```

```

Qy 191 ProGlyValSerThrVal-----ProAsnThrThrGlnAlaSerThr-Pr 205
Db 102 CCTCCACCAACCACTTGATTTGTATTTGTATTCGCGAACCACTCCGCGCAACCGGA 161
Qy 205 opProGlnThrGlnThrPro-----GlnProAsnProProProValGlnAlaThrPr 222
Db 162 TCCCAACCGCACTCCACCAATTTCAACGACGACCGCAATGACCAACACACC 221
Qy 222 ohisProPheProAlaValThrProAspLeuIleValGlnThrProValMetThrValVa 242
Db 222 GCCTCCACCAAGACATCACTCCACTCCAGTAGCGACCAACCA-----CCAGCACT 275
Qy 242 lProProGlnProLeu-----GlnThrProProProValProPr 255
Db 276 TCTCCAAAACCTTTACCGGACCGTGTCTCTCGCAAAACCAACACACACTCTCTCCACC 335
Qy 255 ogInProGlnProProProAlaProAlaProGlnProValGlnSerHisProProIle1 275
Db 336 AGCTATTACCAACCACTCCACCAAGTATTACACCTCCACTATCACTCTCCACACAGC 395
Qy 275 eAlaAlaThrProGlnProValIleThrIleValIleValIleValIleAlaSerThrTh 295
Db 396 TATTACCAACCACTCTCCACTA-----GCGAC 422
Qy 295 rThurProThrThIleAspProIleHisGluProProSerLeuProProGluProIleTh 315
Db 423 CACACACCGGCACTTCTCCAAAACCTTACCGCAACCGTGTCTCT---CCACAAAC 479
Qy 315 rThurIleuGlyGlnArgGlnSerSerArgProValIleProProIleIleValIleVal 335
Db 480 CAGG-----CCACCTCTCCACCAACCAACCAACCAACCAAC 509
Qy 335 lProAspSerGlnGlnHisProAlaProGluIleSerSerIleValIleSerGluGlnLeu 355
Db 510 ACCACTATCACTCT-----CCTCTGTTGGATATATGTTCCAGAAATGATACCAACTAAA 563

```

```

Qy 355 sCysCysSerGlyIleLeuIleGluMetPheAlaIleLysHisAlaIleTyrAlaTrp 375
Db 564 GATATGTGCGGAAATCTGACATTAGTACGGTCTTCAACCACTGGAAGACAGCAAC 623
Qy 375 oPheTyrIleProValAspValGlnAlaLeuGlyLeuHisAspTyrCysAspIleIle 395
Db 624 A-----TGTGTTCCTCAATCCG 641
Qy 395 shiSPrometAspMetSerThrIleLys-----SerLysLeuGlnAlaArgIleLys 412
Db 642 AATGTGCTGATCTTGATGACGACTTGTGTTTGTAAATGCTGAGACACCACTTT 701
Qy 412 rArgAspAlaGlnGluPheGlyAlaAspValArgLeuMetPheSerAsnCysTyrIleTy 432
Db 702 TTCCTCTCTCCCAATTCGGC-----ACTCTTTAAAGTTGT----- 741
Qy 432 rAsnProProAspHisGluValAlaIleAlaIleValIleValIleValIleValIle 452
Db 742 -----GGTCTGAGATTCCCAAGGCTTTAGCTG 770
Qy 452 rArgPheAlaLysMetProAspGluProGluGluProValAlaIleValSerSerProAl 472
Db 771 T-----CCCGGTCCATCAACCAATCTCACCAACACC 803
Qy 472 aValProProProThrLysValValAlaProProSerSerSer 486
Db 804 ACTTCTCCCAAACTTTAAACCGGCTCGGCAACCAACACA 846

```

```

RESULT 15
US-09-764-864-373
; Sequence 373, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: Patencin Ver. 2.0
; SEQ ID NO 373
; LENGTH: 2653
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-764-864-373

```

```

Alignment Scores:
Pred. No.: 2.46e-05 Length: 2653
Score: 225.00 Matches: 136
Percent Similarity: 30.78% Conservative: 81
Best Local Similarity: 19.29% Mismatches: 197
Query Match: 6.18% Indels: 291
DB: 10 Gaps: 23
US-09-700-590A-22 (1-688) x US-09-764-864-373 (1-2653)

```

```

Qy 50 ThrSerAsnProAsnLysProLysArgGlnThr-----AsnGlnLeu 63
Db 293 ACCTCTGATGATGCTGATGCTCAAAAGAAAAGAAAAGAAACATAAGAACGACCAATG 352
Qy 64 GlnTyrLeuLeuArgValIleLysThrIleTyrIleHisGlnPheAlaTrpPhe 83
Db 353 AAAATTTTATTCATGTTGTTCTTGAAGCT-----ValAspAlaValLysLeu 382
Qy 84 GlnGlnPro-----ValAspAlaValLysLeu 92
Db 383 CAGAGACCAAGGTTACAGCAGAAAGCTTTGACCTATTATGTTTAAACATCAACAAAAG 442
Qy 93 AsnLeuProAspTyrTyrIleIleLys-ThrProMetAspMetGlyThrIleLysIly 112
Db 443 GACTATCCGATTTATTAATAATCATCTTGAGCCCAATGAGCTTGAAAATATATTTAGCA 502

```

```

Qy 112 sarGleuGluAsnAsnTyrTyrPAsnAlaGlnGluCysIleGlnSerPheAsnThrMe 132
Db 503 TAAATCGCGAATGACAATATGCTGGTAAGAGGAATGATGAAAGCATGAAGCTCAT 562
Qy 132 tPheThrAsnGlyTyrIleTyrAsnLysProGlyAspSerIleValIleuMetAlaGluAl 152
Db 563 GTTCCGGAATGCCAGGCACTAATGAGAGGGCTCCCGAGTTTAATATGATGACATAT 622
Qy 152 aleuGluLysLeuPheLeuGlnLysIleAsnGluLeuProThrGlnGluThrGluIleMe 172
Db 623 CTGGAGAAAGTACTCAAGAGAAA----- 647
Qy 172 tIleValGlnAlaLysGlyArgGlyArgGlyGlyLysGluThrGlyThrAlaLysProGly 192
Db 648 -----AGGAAAGCGCGGC----- 662
Qy 192 yValSerThrValProAsnThrThrGlnAlaSerThrProProGlnThrGlnThrProGly 212
Db 662 ----- 662
Qy 212 nProAsnProProProValGlnAlaThrProHiserProPheProAlaValThrProAspLe 232
Db 662 ----- 662
Qy 232 uIleValGlnThrProValMetThrValValProProGlnProLeuGlnThrProProPr 252
Db 663 -----CC 664
Qy 252 oValProProGlnProGlnProProProAlaProAlaProGlnProValGlnSerHisPr 272
Db 665 ACTGCTCGATGATGATGAC----- 683
Qy 272 oProIleIleAlaIleThrProGlnProValLysThrLysLysGlyVal-----Ly 289
Db 684 -----ATGCTTCTCCCAAACTCACTGAGTAGAGAGTGCGCATTTCTCTAAAAA 736
Qy 289 sarGlyAlaIleAspThrThrThrProThrThrIleAspProIleGluProProSerLe 309
Db 737 ATCAAATACATGACTCCAAATGCAGCAGAACTAATAGAGCTATGA----- 785
Qy 309 uProProGlnProLysThrThrLysLeuGlyGlnArgIleGlnSerSerArgProValLys 329
Db 786 -GCTGTAAAGACTATAGTATAGAGGGGTGCGCCCTCAGTGCATATTTCTGAGGCT 844
Qy 329 sProProLysAspValProAspSerGlnGlnHisProAlaProGluLysSerSerLys 349
Db 845 TCCTCTAGATCTGAGTGCCT----- 866
Qy 349 sValSerGlnLeuLysCysCysSerGlyIleLeuLysGluMetPheAlaLysLysH 369
Db 866 ----- 866
Qy 369 sAlaAlaTyrAlaTrpProPheTyrLysProValAspValGluAlaLeuGluLysH 389
Db 867 -----GA 868
Qy 389 pTyrCysAspIleIleLysHisProMetAspMetSerThrIleLysSerLysLeuGluAl 409
Db 869 CTACTATCTGACTATTAAGAACCCATGACATGAAAAAATTCAGAAATCATGATGCG 928
Qy 409 aArgGluTyrArgAspAlaGlnGluPheGlyAlaAspValArgLeuMetPheSerAsnCy 429
Db 929 CAACAAGTACCAAGTATATGACTATATGTTGAGAGACTTTGTGATGATGTTTAATATCC 988
Qy 429 sTyrLysTyrAsnProProAsp-----HisGluVal 439
Db 989 CTGTACGTACATAGCGGAGGTCTTTGATCTACAAAGATGCTTTGTTCTACAAAGT 1048
Qy 439 lValAlaMetAlaArgLys----- 445
Db 1049 CTTGCTTGAACACGCGAGAGACTGTGAGAGAGATGAGGACTCTCATGTCCAAATGTGAC 1108

```

```

Qy 446 -----LeuGlnAspValPheGluMetArgPheAlaLysMetProAspGluProGluGly 463
Db 1109 TTTGCTGATTCAAGAGCTTTCACCAATCTTTTGTGTGATGATGATGATGATGATGATGA 1168
Qy 463 uProValAlaValSerSer-----ProAlaValProPro-----Pr 476
Db 1169 TGAAGGAAGATGCTACAGCCATCTTTAGCAGAAATTCCTGCTGTGATGCCAACTTCC 1228
Qy 476 oThrLysValAlaIleProProSerSerSerAspSerSerSerSerSerSerSerAspSe 496
Db 1229 TAACAA-----CCACCCCTTACATTTGACATATTAAGAAATGTTGAAATTA 1279
Qy 496 r-----AspSerSerThrAspAspSerGluLys-----GluArgAlaGly 509
Db 1280 TCGCTACCGTGGCTGATTATTTTCAAGAGCATATGTTGAAGTATTTGAGAACGACG 1339
Qy 509 nArgLeu-----AlaGluLeuGlnGluGlnLe 518
Db 1340 AAGGATGAATCGACAGATTCAGAAATATATGAAGATGCGATGACATTCACGAGTTT 1399
Qy 518 uLysAlaValHisGluGlnLeuAla-----AlaLeuSe 529
Db 1400 TATTAAATTCGTGATGAATCTGCAAAAATGAGAGATTTCTTCCACCGGACTCAG 1459
Qy 529 rGlnProGlnGlnAsnLysProLysLysLysGluLysAspLysLysGluLys----- 546
Db 1460 CTATACCACAAAACATTTGATATGATGATGAGAAAGAGAGAAAGGAAATTTGCCAA 1519
Qy 547 -----LysLysGluLysHisLysArgLysGluGluValGlnGluAsnLysLysSerLys 564
Db 1520 AGAAATGAGAGAGATTAACCTAAACGAGAAAGAAAGAAAGAAAGAAAGAAAGATGA 1579
Qy 564 sAlaLysGluProProProLysThrLysLysAsnAsnSerSerSerSerSerValSe 584
Db 1580 AGATTCTCTGTGCTGTGATGAGGCTCTCAGGCTTACATGCAATACAGCCAGACTGTAG 1639
Qy 584 rLysLys-----GluProAlaProMetLys 592
Db 1640 CTTTAAAAACGACATGTACATGTGGAGATTAAGTATGTGTGAACCTCGAGAGCCAA 1699
Qy 592 sSerLysProProProThr----- 598
Db 1700 CTTAACAACACATATCGTCTGTATTTGAAGAAGCTGGAGAGATTCAGCTGTGAATAATG 1759
Qy 599 -----TyrGlnSerGlnGluGluAspLysCysLysProMetSerLys 612
Db 1760 GTTGTATGCGTGTGTTTACGAGCAAAATGAATTCACCTGCTACACGAAATTT 1819
Qy 612 rGlnGluLysArgGlnLeuSerLeuAsp-----IleAsnLysLeuProGlyGluLys---Le 630
Db 1820 TCTAGAAAAGAAAGTCTTTTAAGAGTGACTATTCACAAAGTTTCAGTTAGTAAATTTCT 1879
Qy 630 uGlyArgValVal 634
Db 1880 AGGCAAGTGTGTG 1892

```

Search completed: March 14, 2003, 22:45:21
 Job time : 208 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 14, 2003, 18:46:26 ; Search time 2111 Seconds
(without alignments)
5278.302 Million cell updates/sec

Title: US-09-700-590A-22
Perfect score: 3639
Sequence: 1 MSASGPGTRLRLNLPVMDG.....ALCHLLFAEKETFKRLKLM 688

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 809774376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DBV=xlp
-Q/cgn2.1/USFTO.spool/US09700590/runat.14032003.140630.3244/arp.query.fasta.1.839
-DB=EST -OPMT=fastap -SUFFIX=1st -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsun62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pcp -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pcp -NOR=ext -HEASize=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09700590.cgn2.1.1.2874.arp.runat.14032003.140630.3244 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEORDER -NEG_SCORES=0 -WAIT -LONGLOG -DEXTIMEOUT=120
-WARN_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7
-YGAPOP=10 -YGAPEXT=0.5 -DELop=6 -DELEXT=7

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_estl:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pin:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_tod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	* Query Match	Length	DB	ID	Description
1	2914	80.1	1716	11	BC000156	BC000156 Homo sapi
2	2898	79.6	1743	11	BC030158	BC030158 Homo sapi
3	2814.5	77.3	1776	11	BC011541	BC011541 Mus muscu
4	1919.5	52.7	2085	11	BC010699	BC010699 Homo sapi
5	1590.5	43.7	1641	11	BC004832	BC004832 Mus muscu
6	1535.5	42.2	2359	11	BC007715	BC007715 Homo sapi
7	1494	41.1	987	14	BC051563	BC051563 AGENCOURT
8	1484.5	40.8	2440	11	BC001885	BC001885 Homo sapi
9	1436	39.5	893	14	BC021977	BC021977 AGENCOURT
10	1408	38.7	933	14	BC021440	BC021440 AGENCOURT
11	1408	38.7	951	14	BC046784	BC046784 AGENCOURT
12	1357.5	37.3	925	14	BC045818	BC045818 AGENCOURT
13	1352	37.2	975	14	BC065318	BC065318 AGENCOURT
14	1336.5	36.7	815	14	BC074932	BC074932 UI-M-FCO-
15	1319.5	36.3	926	14	BC050970	BC050970 AGENCOURT
16	1315	36.1	919	14	BC052326	BC052326 AGENCOURT
17	1311.5	36.0	931	14	BC050713	BC050713 AGENCOURT
18	1311	36.0	899	14	BC047311	BC047311 AGENCOURT
19	1308	35.9	938	14	BC054353	BC054353 AGENCOURT
20	1280	35.2	740	14	BC068304	BC068304 UI-CF-EN1
21	1284	34.2	1106	13	BM452148	BM452148 AGENCOURT
22	1225.5	33.7	971	14	BC052321	BC052321 AGENCOURT
23	1201	33.0	1452	11	BC028126	BC028126 Homo sapi
24	1200	33.0	1538	11	BC017582	BC017582 Homo sapi
25	1195	32.8	1470	11	BC005281	BC005281 Homo sapi
26	1150.5	31.6	919	12	BF180938	BF180938 601805044
27	1149	31.6	1011	11	BC013666	BC013666 Mus muscu
28	1129	31.0	709	13	BF141784	BF141784 602911186
29	1126	30.9	675	12	BF550759	BF550759 UI-R-CO-J
30	1126	30.9	1055	12	BC082432	BC082432 602727345
31	1114	30.6	641	13	BF190731	BF190731 603065284
32	1099	29.2	658	13	BM389207	BM389207 UI-R-DDO-
33	1072	29.5	682	13	BF125263	BF125263 602952637
34	1066	29.3	621	10	AM539257	AM539257 CO117F08-
35	1061	29.2	631	12	BF712714	BF712714 MI-P-H3-a
36	1053	28.9	627	14	BM679991	BM679991 UI-E-EOL-
37	1052	28.9	666	12	BF718610	BF718610 KESTR1 no
38	1039.5	28.6	786	13	BF185291	BF185291 603379265
39	1038	28.5	610	10	AM142274	AM142274 EST92510
40	1036	28.5	619	14	BO939643	BO939643 AGENCOURT
41	1027	28.2	789	13	BC963832	BC963832 602831494
42	1018	28.0	589	13	BM226660	BM226660 K0231D10-
43	1018	28.0	615	12	BC931946	BC931946 602409863
44	1002.5	27.5	629	9	AI552951	AI552951 Y661901.Y
45	992	27.3	902	14	BC023237	BC023237 AGENCOURT

ALIGNMENTS

RESULT 1
LOCUS BC000156
DEFINITION Homo sapiens, Similar to bromodomain-containing 4, clone
ACCESSION BC000156
VERSION BC000156.1 GI:14704448
KEYWORDS
SOURCE
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
AUTHORS Strausberg, R.
TITLE Direct Submission

JOURNAL

Submitted (03-NOV-2000) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK
COMMENT

NIH-MGC Project url: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ULNL)
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: villalon@bcm.tmc.edu
Villalon, D.K., Luna, R.A., Hale, S.M., Hulik, S., Lu, X., Garcia, A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W., Muzny, D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ULNL at: <http://image.llnl.gov>
Series: IRAC Plate: 3 Row: e Column: 7
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 3115203
This clone has the following problem: frame shifted.

FEATURES

source

1. 1716
Location/Qualifiers

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2900081"
/tissue_type="Placenta, choriocarcinoma"
/clone_id="NIH MGC_10"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"

BASE COUNT 474 a 549 c 427 g 266 t
ORIGIN

Alignment Scores:

Pred. No.: 1,77e-160 Length: 1716
Score: 2914.00 Matches: 548
Percent Similarity: 100.00% Conserved: 1
Best Local Similarity: 99.82% Mismatches: 0
Query Match: 80.08% Indels: 0
Gaps: 0

US-09-700-590a-22 (1-688) x BC000156 (1-1716)

QY 1 MetSerAlaGluSerGlyProGlyThrArgLeuArgAsnLeuProValMetGlyAspGly 20
Db 69 ATGCTGGGAGAGAGCGCCCTGGAGCATTTGAGAAATCTGCCGATATGGGGATGGA 128
QY 21 LeuGluThrSerGlnMetSerThrThrGlnAlaGlnGlnProGlnProAlaAsnAla 40
Db 129 CTAAAGAACTTCCAAATGTTCAACACAGAGCCGACCCCAAGCCAGCCAAAGCGCA 188
QY 41 AlaSerThrAsnProProProGluThrSerAsnProAsnLysProLysArgGlnThr 60
Db 189 GCCAGCACCACCCCGCCCGCCAGAGACCTCCAACTTAACAGCCCAAGAGCGAGACC 248
QY 61 AsnGlnLeuGlnTyrLeuLeuArgValValLeuLysThrLeuTyrLysHisGlnPheAla 80
Db 249 AACCAACTGCAATACCTGCTCAGAGTGTGCTCAAGACACTATGAAACACAGTTTCA 308
QY 81 TyrProPheGlnGlnProValAspAlaValLysLeuAsnLeuProAspTyrTyrLysIle 100
Db 309 TGGCTCTTCCAGACGCTGTGATGCCGTCAAGCTGAACCTCCCTGATTAATAAGATC 368
QY 101 IleLysThrProMetAspMetGlyThrIleLysLysArgLeuGlnAsnAsnTyrTyrTrp 120
Db 369 ATTAAAGCGCTTATGATATGGAGCAATTAAGAGCGCTTGAAAAACAATTAATCTGG 428
QY 121 AsnAlaGlnGluCysIleGlnAspPheAsnThrMetPheThrAsnCysTyrIleTyrAsn 140

Db 429 AATCTCAGAGATGATATCCAGGACTTCAACAATGTTTACAAATGTTATCATTTACAC 488
QY 141 LysProGluAspAspIleValLeuMetAlaGlnAlaLeuGluLysLeuPheLeuGlnLys 160
Db 489 AAGCTGTGAGATGACATAGTCTTTAATGGCAGAGCTGTGGAAAAGCTCTTCTTCAAAA 548
QY 161 IleAsnGluLeuProThrGluGluThrGluIleMetIleValGlnAlaLysGlyArgGly 180
Db 549 ATAAATGAGTACCAACAGAAAGAAACGAGATCATGTATGCTCCGAGAAAAGAGAGGA 608
QY 181 ArgGlyArgLysGluThrGlnTyrThrAlaLysProGlyLysValSerThrValProAsnThr 200
Db 609 CGTGGAGAAAGAAACAGAGGACGACAAACCTGGCTTTCCAGGTACCAACACAACT 668
QY 201 GlnAlaSerThrProProGlnThrGlnThrProGlnProAsnProProProValGlnAla 220
Db 669 CAAGCATGACTCTCTCCGAGACCCCAACCCCTCAGCGAATCTCTCTGTGTGAGGCC 728
QY 221 ThrProHisProPheProAlaValThrProAspLeuIleValGlnThrProValMetThr 240
Db 729 ACGCTCACCCCTTCCCTGCGGTCAACCCCGACCTCATGCTCAGACCCCTGTATGACA 788
QY 241 ValValProProGlnProProLeuGlnThrProProProValProProGlnProProPro 260
Db 789 GTGTGTGCTTCCCGACGACATGACAGACCCCGCCAGTGTGCCCCAGCAACACCCCA 848
QY 261 ProAlaProAlaProGlnProProValGlnSerHisProProIleIleAlaAlaThrProGln 280
Db 849 CCCCTCAGTCTCCCGACGCCGTACAGACCCACCCACCATCATCGGGCCACCCACAG 908
QY 281 ProValLysThrLysLysGlyValLysArgLysAlaAspThrThrThrProThrThrIle 300
Db 909 CCTGTGAAGACAAAGAGGAGGTGAAGAGAAAGCAACACACACCCACACCATTT 968
QY 301 AspProIleHisGluProProProSerLeuProProGluProLysThrThrLysLeuGln 320
Db 969 GACCCCATTCACGAGCACCTGCTGCTCCCGAGGCCAAGACACCAAGCTGGGACAG 1028
QY 321 ArgArgLysSerSerArgProValLysProProLysLysAspValProAspSerGlnGln 340
Db 1029 CGGCGGAGAGACACCCGCTGTGAATCTCCAAAGAGACGTCGCCGACTCCACAG 1088
QY 341 HisProAlaProGluLysSerSerLysValSerGlnLeuLysCysCysSerGlyIle 360
Db 1089 CACCAGACACAGAGAGAGAGCAAGGTCTCGGAGCAGCTCAAGTCTGACGCGCATC 1148
QY 361 LeuLysGluMetPheAlaLysLysHisAlaAlaTyrAlaTrpProPheTyrLysProVal 380
Db 1149 CTCAAGAGAGATGTTGCCAAGAAACAGCGCTTACGCTGGCTTCTACAAAGCTGTG 1208
QY 381 AspValGluAlaLeuGlyLeuHisAspTyrCysAspIleIleLysHisProMetAspMet 400
Db 1209 GACGTGAGGCACTGGGCTTACAGCACTACTGTACATCATCAAGCACCCCATGAGCATG 1268
QY 401 SerThrIleLysSerLysLeuGlnAlaArgGluTyrArgAspAlaGlnGluPheGlyAla 420
Db 1269 AGCAATCAAGTCTTAATCTGAGAGCCCGTAGTACCGTATGCTCAGAGATTGGTGT 1328
QY 421 AspValArgLeuMetPheSerAsnCysTyrLysTyrAsnProProAspHisGlnValVal 440
Db 1329 GACGTCCGATTGATGTTCTCCAACTGCTATAGTAAACCTCTGTGACATGAGAGTGTG 1388
QY 441 AlaMetAlaArgLysLeuGlnAspValPheGluMetArgPheAlaLysMetProAspGlu 460
Db 1389 GCCATGGCCGCAAGCTCCAGAGATGTGTGAATAGCGCTTTCAGAGATGCCGAGAGAG 1448
QY 461 ProGluGluProValValAlaValSerSerProAlaValProProProThrLysValVal 480
Db 1449 CCTGAGAGAGCAGTGTGTGCGGTCTCTCCCGGCAATGCCCCCTCCACCAAGTTGTG 1508
QY 481 AlaProProSerSerSerAspSerSerSerSerSerSerSerSerSerSerSerThr 500
Db 1509 GCCCCGCTCATCTCAGAGAGAGAGAGAGATGCTCTCGAGCAGTGAAGTTGACT 1568

OY	501	AAGAAATGGTCTGTAACAGAACAAGTTAATAATTGAATCAACTCCTTCAGCGAGCCGTCAAAGGCC	
Dn	1659	GATGCATCTTGAGAGAGAGAACACCAGCGCGCTGAGCTCAACTCCAAGAGCATCTCAAAGCC	1628
OY	521	VAlHtSGlUGlnLeuAlALaleUSeRgInPrOGlngInAsnlYaProlYeLySLySglU	540
Dn	1629	GTGCACGAGCAGCTTGACGCCCTCTCTCACGCCCCAGCAAAACAAAAGAAAAAGAG	1688
OY	541	LysApplYelYLylGLlyULyLylSlLU	549
Dn	1689	AAAGCAAGAGGAAAAAAAAAAAAA	1715
RESULT 2			
BC030158			
LOCUS			
DEFINITION	BC030158	1743 bp mRNA linear HTC 20-MAY-2002	
	Homo sapiens, Similar to Bromodomain-containing 4, clone IMAGE:5086929, mRNA.		
ACCESSION	BC030158		
VERSION	BC030158.1	GI:20988807	
KEYWORDS	HTC.		
SOURCE	Homo sapiens.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1 (bases 1 to 1743)		
AUTHORS	Straubeberg,R.		
TITLE	Direct Submission		
JOURNAL	Submitted (07-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,		
REMARK	USA NIH-MGC Project URL: http://mgc.nci.nih.gov		
COMMENT	Contact: MGC help desk		

Email: c9apbbs1@mail.nih.gov
Tissue Procurement: CGAP (Stanford)
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland!
Web site: <http://www.nisc.nih.gov/>
Contact: nisc_mgc@nhgri.nih.gov
Ahter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
Blakesley, R.W., Boulter, G.G., Breen, K., Brinkley, C., Brooks, S.,
Dietrich, N.L., Grantle, S., Guan, X., Gupta, D., Haghighi, P.,
Hansen, N., Ho, S.-L., Karlins, B., Latic, P., Legspsy, R., Maduro, O.L.,
Mastello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J.,
Pearson, R., Stantirrop, S., Thomas, P.J., Touchman, J.W., Tsurgon, C.,
Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A.,
Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/ILNL at: <http://image.llnl.gov>
Series: IRAL, Plate: 39 Row: n Column: 1
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 7657217
This clone has the following problem: frame shifted.

FEATURES	Location/Qualifiers
source	1. .1743

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5086929"
/tissue_type="liver, hepatocellular carcinoma"
/clone_id="NH1_MGC_100"
/lab_host="DH10B"
/host="Vector: pOTB7"

```

Alignment Scores:	
Pred. No.:	1.53e-159
Score:	2898.00
Length:	1743
Matches:	545

Percent Similarity:	99.82%	Conservative:	2
Best Local Similarity:	99.45%	Mismatches:	1
Query Match:	79.64%	Indels:	0
DB:	11	Gaps:	0

Qy	1	MetSerAlaGluSerGlyProGlyThrArgLeuArgAsnLeuProValMetGlyAspGly	20
Db	99	ATGTCGTGGAGAGACGGGGCCCTGGGACGAGATTGAGAAATCTGCCAGTAAATGGGGAGATGGA	158
Qy	21	LeuGlnThrSerGlnMetSerThrThrGlnAlaGlnAlaGlnProGlnProAlaAsnAla	40
Db	159	CTAGAAACTTCCCAATATGTCTACAAACAGAGCCAGGCCCAACCCCGACGACCAAGCA	218
Qy	41	AlaSerThrAsnProProProProGlnThrSerAsnProAsnGlyProLysArgGlnThr	60
Db	219	GCACGACCAACCCCCCGCCGCCAGACCTCCCAACCTTACAAAGGCCCAAGAGGCAACC	278
Qy	61	AsnGlnLeuGlnTyrLeuLeuArgValValLeuValSerThrLeuTyrLysHisGlnPheAla	80
Db	279	AACCAACTGCATACCTGCTCAGAGTGGTGCTCAAGACACTATGGAAGAACCAGTTTGA	338
Qy	81	TrpProPheGlnGlnProValAspAlaValLysLeuAsnLeuProAspTyrTyrLysIle	100
Db	339	TGGCCTTTCCAGCAGACCTGTGTGAAGCCGTCAACGCTGAACCTCCCTGATTCTATACATC	398
Qy	101	IleLysThrProMetAspMetGlyThrIleLysLysArgLeuGlnAsnAsnTyrTyrTrp	120
Db	399	ATTAAAGCCCTATAGGATATGGAGAACATAAAGAGCGCTTGGAAAACAATATTACTGG	458
Qy	121	AsnAlaGlnGlnCysIleGlnAspPheAsnThrMetPheThrAsnGlyTyrIleTyrAsn	140
Db	459	AATGCTCAGGAATGATGTCACAGACTTCAACACTATGTTTCAAAATTTTTCATCTAACAC	518
Qy	141	LysProGlyAspAspIleValLeuMetAlaGlnAlaLeuGlnLysLeuPheLeuGlnLys	160
Db	519	AAGCCTGGAGATGACATATGTTTAAATGGCAGAGACCTGTGAAAAAGCTTCTTGGCAAAA	578
Qy	161	IleAsnGluLeuProThrGlnGlnThrGlnIleMetIleValGlnAlaLysGlyArgGly	180
Db	579	ATAAATGAGCTAACCCACAGAAAGAACCGAGATCTAGTATGCCAGGCAAAAGGAGGGA	638
Qy	181	ArgGlyArgLysGlnThrGlyThrAlaLysProGlyValSerThrValProAsnThrThr	200
Db	639	CGTGGAGAGAAAGAAACAGGAGACGAAACCTGGCTTTCCACGATACCAACACAACT	698
Qy	201	GlnAlaSerThrProProGlnThrGlnThrProGlnProAsnProProProValGlnAla	220
Db	699	CAAGCATGACCTCTCCGCAACCCAGACCCCTCAGCCGGAATCCTCTCTGTGCAGGCC	758
Qy	221	ThrProHisProPheProAlaValThrProAsnLeuIleValGlnThrProValMetThr	240
Db	759	ACGCTCAACCCCTTCCCTGCGCGTACCCCGGACCTCATGCTCCAGACCCCTGTATACA	818
Qy	241	ValValProProGlnProLeuGlnThrProProProValProProGlnProGlnProPro	260
Db	819	GTGGTGCTCCCCAGACCACTGCAACGCCCGCCGCACTGCCCCCGCCAGCCACAACCCCA	878
Qy	261	ProAlaProAlaProGlnProValGlnSerHisProProIleIleAlaAlaThrProGln	280
Db	879	CCCGCTCCAGCTCCCGACGCCGTGACAGGCCACACCCATCATCGGGCCACCCCAAG	938
Qy	281	ProValLysThrLysLysGlyValLysArgLysAlaAspThrThrThrProThrThrIle	300
Db	939	CTGTGAAGACAAAGAGGGAGTAAAGGAGAAACAGACACACACCCCAACCACTT	998
Qy	301	AspProIleHisGluProProSerLeuProProGlnProLysThrThrLysLeuGlyGln	320
Db	999	GACCCCATTCACAGACCACTCGCTGCCCCCGGAGCCAAAGACCAACAGCTGGCCAG	1058
Qy	321	ArgArgGluSerSerArgProValLysProLysLysAspValProAspSerGlnGln	340

QY 64 GlnTyrLeuLeuValValLeuYsThrLeuTyrLysHisGlnPheAlaTyrProPhe 83
 |||||
 Db 276 CAGTACATGAGATGTGTGTGTGAAGACGCTGGAACACCAAGTTGCGCTGGCCCTTC 335
 QY 84 GlnGlnProValAspAlaValLeuLeuAsnLeuProAspTyrTyrLysHisLeuLeuThr 103
 |||||
 Db 336 TACACGCCCGGAGCGCAATGAACTGACCGGATTCATTAATAATATTAATAAAC 395
 QY 104 ProMetAspMetGlyThrTyrLeuLysArgLeuGlnAsnAsnTyrTyrTyrAsnAlaGln 123
 |||||
 Db 396 CCAATGATATGAGGAGCATTTAAGAAGAGACTAAGAAATATATTTATTTGAGTGCAAGC 455
 QY 124 GluCysIleGlnAspPheAsnThrMetPheThrAsnCysTyrIleTyrAsnLysProGly 143
 |||||
 Db 456 GAATGTATGACGAGCTTCAACACCATGTTTCAAAATGTTTCAATTATTAACAGCCCA 515
 QY 144 AspAspIleValLeuMetAlaGlnAlaLeuGlnLysLeuPheLeuGlnIleAsnGln 163
 |||||
 Db 516 GATACATATAGCTAATAGCCCAAGCTTTAGAGAAATTTTTCACAAATAATGAGCCGAG 575
 QY 164 LeuProThrGlnGluThrGluIleMetIleValGlnAlaLysGlyArgGlyArg----- 181
 |||||
 Db 576 ATGCCCAAGAGAGAGTTGAATTTATTTACCCCTGCTCCAAAGCGCAAGCTCGAAGCG 635
 QY 182 -----GlyArgLysGluThrGlyTyrThraLysProGlyValSerThrValProAsnThr 199
 |||||
 Db 636 GCTCGGAGCGCCAGAGCGCAGGTACAGCA-----GTGGCGCGCGTGTCTCTCTC 689
 QY 200 ThrGlnAlaSerThrProProGlnThrGlnThrPro-----GlnProAsnProProAla 218
 |||||
 Db 690 TCCCGAGG-----ACCCCTTTTCAGAGCGTGGCCCGCCACCTCTC 728
 QY 219 GlnAlaThrProHisProPheProAlaValThrProAspLeuIleValGlnThrProVal 238
 |||||
 Db 729 TCCAGAGAGCGCC-----GTCTGCTGCGCAACCCCTCTA 761
 QY 239 MetThrValValProProGlnProLeuGlnThrProProProValProProGlnProGln 258
 |||||
 Db 762 CCAACCATCATCGCAAC-----GTACAGTGTGTCCAGTCCCGCCAGCTGCCCGC 812
 QY 259 ProProProAlaProAlaProGlnProValGlnSerHisProProIleIleAlaIleThr 278
 |||||
 Db 813 CCACCTCTCT-----CTGCACACCCCATGTCCTCGTGTCTCTCTC 857
 QY 279 ProGlnProValLysThrLysLysGlyValLysArgLysAlaAspThrThrThrProThr 298
 |||||
 Db 858 CCG-----CTGTCTGTCAGAAAGAGGCGGTGAAGCGAAAGCAGACACCATCTCCACG 914
 QY 299 -----ThrIleAspProIleHisGlnProProSerLeuProProGlnProLys 314
 |||||
 Db 915 ACGTGGCGCATCATCGCAGCGGAGGTAGTGTGCGCCCGCGCTG---TCAGACCCCAAG 971
 QY 315 ThrThrLysLeuGlyGlnArgArgLysSer---SerArgProValLysProProLysLys 333
 |||||
 Db 972 CAGCGCAAAAGTGTGGCGCGGCGGAGAGTGTGTGCGCCCATCAACCTCTCCCAAGAG 1031
 QY 334 AspValProAspSerGlnGlnHisProAlaProGlnLysSerSerLysValSerGln 353
 |||||
 Db 1032 GACCTGGAGAGAGCGGAGGTGCGCCAGCACGAGCAGCAAGAGGAGAGGTGTCGAGCAC 1091
 QY 354 LeuLysCysCysSerGlyIleLeuLysGlnMetPheAlaLysLysHisAlaAlaTyrAla 373
 |||||
 Db 1092 CTACGCTACTGAGACAGCATCTCTAGGAGATGCTATCAAAACACACGCGGCTTACGCC 1151
 QY 374 TyrProPheTyrLysProValAspValGlnAlaLeuGlyLysHisAspTyrCysAspIle 393
 |||||
 Db 1152 TGGCCCTTCTCAAGCCAGTGTGATGCGCGCGCTGTGAGCTCACAGACTACACACACATTC 1211
 QY 394 IleLysHisProMetAspMetSerThrIleLysSerLysLeuGlnAlaArgLysTyrArg 413
 |||||
 Db 1212 ATCAAGACCCGATGAGCTTACGACCGGTGAAGAGAGATGATGCGGAGATGCCA 1271

QY 414 AspAlaGlnGluPheGlyAlaAspValArgLeuMetPheSerAsnCysTyrLysTyrAsn 433
 |||||
 Db 1272 GAGCCACAGGCGCTTGTGCTGATGTCGGCTCATGTCTCGAATTGCTCAAAATCAAT 1331
 QY 434 ProProAspHisGlnValAlaAlaMetAlaArgLysLeuGlnAspValPheGluMetArg 453
 |||||
 Db 1332 CCCCACACACAGAGGTGTGGCCATGCGCCGGAAGCTCCAGAGCGTGTGTGAGATGAG 1391
 QY 454 PheAlaLysMetProAspGluProGluGluProValAlaAlaValSerSerProAlaVal 473
 |||||
 Db 1392 TTGGCCAAAGATGCGCAGATGAGCC-----GTGAGCGCACCGCGGCTG 1433
 QY 474 ProProThrThrLysValAlaAlaProProSerSerSerAspSerSerSerSer 493
 |||||
 Db 1434 CTGCCCCCGCGCGCCCATATGTGAGCAGAGCGGCTGTAGACGACCCCTTACAGTGAAG 1493
 QY 494 SerAspSerAspSerSerThrAspAspSerGluGlnGluArgAlaGlnArgLeuAlaGln 513
 |||||
 Db 1494 AGCTCTTGAGCTTACGACACTCGGACTCGGAGAGAGAGCGGCGCACAGCTGTGCGAG 1553
 QY 514 LeuGlnGlnGlnLeuLysAlaValHisGlnGlnLeuAlaAlaLeuSerGlnProGlnGln 533
 |||||
 Db 1554 CTGCAAGAGCAGCTGAAGCGCGGTGACAGACAGCTGCGCCCTGTCTCAGGCCCGCA 1613
 QY 534 AsnLysProLysLys-----LysGlnLysAspLysLysGluLysLys 547
 |||||
 Db 1614 AACCAACCAAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1673
 QY 548 LysGlnLysHisLysArgLysGlnGlnValGlnGlnLysLysSerLys----- 564
 |||||
 Db 1674 AAGGAG 1733
 QY 565 ---AlaLysGluProProProLysLysThrLysLysAsnSerSerSerAsnVal 583
 |||||
 Db 1734 CCTGCCAAGCAGGCTCAG 1793
 QY 584 SerLysLysGluProAlaProMetLysSerLysProProProThrTyrGlnSerGlnGlu 603
 |||||
 Db 1794 GCCCGAGACAGCTGAAGAAAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1853
 QY 604 GluAspLysCysLysProMetSerTyrGlnGlnLysArgLysLeuSerLeuAspIleAsn 623
 |||||
 Db 1854 GAGAGAGAGAGCGCTCCCATATGAGCTATGAGAAAGCGCAGCTTACCTCGGACATCAC 1913
 QY 624 LysLeuProGlyGlnLysLeuGlyArgValHisIleIleGlnSerArgGluProSer 643
 |||||
 Db 1914 CGGCTGCCCGGAGAGAGAGCTGGCGCGGTAGTGCATCTCCAAATCTCGGAGCCCTCG 1973
 QY 644 LeuLysAsnSerAsnProAspGlnIleGlnIleAspPheGluThrLeuLysProSerThr 663
 |||||
 Db 1974 CTCAAGAGCTTCAACCCCGACGAGATGAATTAATTTGAGACTTGAAGACCACT 2033
 QY 664 LeuArgGlnLeuGlyAlaLeuCysHisLeuLeuPheAlaGlnLysGlu 680
 |||||
 Db 2034 TTGGCGGAACCT-GGAGAGATATGTCAAGTCTTTTCAAAAAAATAAAAAA 2083
 RESULT 5
 BC004832 1641 bp mRNA linear HTC 07-AUG-2002
 LOCUS BC004832
 DEFINITION Mus musculus, similar to RIKEN cDNA 2410084F24 gene, clone
 IMAGE:2650994, mRNA.
 ACCESSION BC004832
 VERSION BC004832.1 GI:14709507
 KEYWORDS
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 1641)
 AUTHORS Strausberg, R.
 TITLE Direct Submission
 JOURNAL Submitted (21-MAR-2001) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer


```

Db 1398 -----GCACCTGCGCTGCGCTCCCGCCAGCGCCCATCTGAGCAG 1439
Qy 484 SerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 503
Db 1440 GGGCTGAGAGACCGGACGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1499
Qy 504 GUGUGUGUAGUAGUAGUAGUAGUAGUAGUAGUAGUAGUAGUAGUAGUAGU 523
Db 1500 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1559
Qy 524 GUGUGUGUAGUAGUAGUAGUAGUAGUAGUAGUAGUAGUAGUAGUAGUAGU 543
Db 1560 CAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1616
Qy 544 LysGluLysLysLysGluLys 550
Db 1617 AAGGAAAAAGAAAAA 1637

RESULT 6
BC007715 2359 bp mRNA linear HTC 12-JUN-2001
LOCUS Homo sapiens, bromodomain-containing 2, clone IMAGE:4302993, mRNA.
DEFINITION BC007715 GI:14712779
ACCESSION BC007715.1 GI:14712779
VERSION 1
KEYWORDS Homo sapiens.
SOURCE Homo sapiens.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2359)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (11-MAY-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc.mgc@hgrl.nih.gov
Shevchenko,Y., Wetherby,K.D., Beckstrom-Sternberg,S.M.,
Benjamin,B., Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S.,
Dietrich,N.L., Guan,X., Gupta,J., Ho,S.-L., Karlins,B., Legaspi,R.,
Lim,M., Maduro,O.L., Mastello,C., Mastrian,S.D., McCloskey,J.C.,
McDowell,J., Pearson,R., Snyder,B., Statistpop,S., Thomas,P.J.,
Tiongson,E.E., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A.,
Zhang,L.-H. and Green,E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/ILNL at: http://image.llnl.gov
Series: IRAL Plate: 17 Row: 0 Column: 3
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 12408641
This clone has the following problem: incomplete processing.
location/Qualifiers
1..2359
/organism="Homo sapiens"
/db_xref="locusID:6046"
/db_xref="taxon:9606"
/clone="IMAGE:4302993"
/tissue_type="uterus", leiomyosarcoma"
/clone_id="NIH_MGC_46"
/lab_host="DH10B-R"
/note="Vector: pOTB7"

BASE COUNT 603 a 683 c 593 g 480 t
ORIGIN

```

```

Alignment Scores:
Pred. No.: 4,24e-80 Length: 2359
Score: 1535.50 Matches: 319
Percent Similarity: 66.61% Conservative: 62
Best Local Similarity: 55.77% Mismatches: 120
Query Match: 42.20% Indels: 71
DB: 11 Gaps: 13

US-09-700-590a-22 (1-688) x BC007715 (1-2359)
Qy 5 SerGlyProGlyThrArgLeuArgAsnLeuProValMetGlyAspGlyLeuGluThrSer 24
Db 770 GCAGACACCAAGGAAAGATTCGAAACCTCTCTCTGATATAGAGGCTTGAGAGGCC 829
Qy 25 GlnMetSerThrThrGlnAlaGlnAlaGlnProGlnProAlaAsnAlaAsnThrAsn 44
Db 830 ACATGGCTTCGGTGGCTGCT--TTGCACTTAACCCCTGCC-----AAC 871
Qy 45 ProProProProGluThrSerAsnProAsnLysProLysArgGlnThrAsnGlnLeuGln 64
Db 872 CCACCAACCCCGGAGGTGTCATCCCAAAAGCAGAGAGAGTTACCAACCAAGCTGCAA 931
Qy 65 TyrLeuLeuArgValValLeuLysThrLeuTyrPheLysHisGlnPheAlaTTPProPheGln 84
Db 932 TACTTACCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 991
Qy 85 GlnProValAspAlaValLysLeuAsnLeuProAspTyrTyrLysIleIleLysThrPro 104
Db 992 CAGCTGTGATGCTGTGCAAACTGGTCTACCGGATTAACAAATTAATAAACAACCT 1051
Qy 105 MetAspMetGlyThrIleLysLysArgLeuGlnAsnLysTyrTyrTyrPheAlaGlnGlu 124
Db 1052 ATGACACATGGTATCTATTAAAGAGAGAGCTGAAACAAATTAATTAATGGCTTCAGAG 1111
Qy 125 CysIleGlnAspPheAsnThrMetPheThrAsnGlyTyrIleTyrAsnLysProGlyAsp 144
Db 1112 TGATGCAAGTTTAAATACATGTTTCAACACTGTTTACATTACCAACCAAGCCACATG 1171
Qy 145 AspIleValLeuMetAlaGlnAlaLeuGlnLysLeuPheLeuGlnLysIleAsnGlnLeu 164
Db 1172 GATTTGTCCTAATGACCAACAAGCTGGAAGAATATTCACAGAGAGTTGCATCATG 1231
Qy 165 ProThrGluGluThrGluIleMetIleValGlnAlaLysGly-----ArgGly--- 180
Db 1232 CCACAGAGAGAAACAGAGCTGTGATGATGATGATGATGATGATGATGATGATGATG 1291
Qy 181 -----ArgGlyArgLysGluThrGlyThrAlaLysProGlyValSerThr 195
Db 1292 AAGTTGGACGCTCCAGGAGAGTTCACAGTCCATCAGGCTGCTGCTGCTTCTTCT 1351
Qy 196 ValProAsnThrThrGlnAlaSerThrProProGlnThrGlnThrProGlnProAsnPro 215
Db 1352 GTGCACACACAGCCCTGTATCTCTCCACCTGAGATA-----CCT 1393
Qy 216 ProProValGlnAlaThrProHisProPheProAlaValThrProAspLeuIleValGln 235
Db 1394 ACCACTGCTCCCAACATTCCTCCCA-----TCAGTATTCC 1432
Qy 236 ThrProValMetThrValValProProGlnProLeuGlnThrProProProValProPro 255
Db 1433 TCTCCACTCTCAAGTCTTGCAC-----TCTGCTGACCCCGCTCTTCTGCT 1480
Qy 256 GlnProGlnProProProAlaProAlaProGlnProValGlnSerHisProProIleIle 275
Db 1481 GTTACTGAGCTCTCTCAGC----- 1501
Qy 276 AlaAlaThrProGlnProValLysThrLysLysGlyValLysArgLysAlaAspThrThr 295
Db 1502 -----CAGCCCTTGCCAGAGAAAAAGCGCTAAAGCGAGAAAGCATATACAC 1549
Qy 296 ThrProThrThrIle-----AspProIleHisGlnProProSerLeu 309

```



```

Db 1550 ACCCTACACCTACAGCATCTTGCGCTCGTTCCTCCACTAGCCCTCCTGGAGCTT 1609
Qy 310 ProProGluProLyThrThrLySleu--GlyGlnArgArgLysSerArgProVal 338
Db 1610 -----GAGCTTAAGCAGACGCGCTTCCCCCTATGCTAGAGAGAGTGTGCCCCCATC 1663
Qy -329 LysProProLySlyAspValProAspSerGlnGlnHisProAlaProGluLySerSer 348
Db 1664 AACCCCCACAGCAAAACCTTGCTGACTCCACACACACACAGAGCTTAAGAAAGA 1723
Qy 349 LysValSerGluGlnLeuLySlyCysSerGlyYlleuLySglMetPheAlaLySly 368
Db 1724 AGCTTTCAGAACAGTTTAAACATTGCAATGGCATTTTGAAGAGATTACTCTTAAGAG 1783
Qy 369 HisAlaAlaTyrAlaTyrProPheTyrLySProValAspValAlaLeuGlyLeuHis 388
Db 1784 CAGTGGCTCATGCTTGCGCTTCTTATTAACCAAGGATGATCTTGTGCACTTGCGCAT 1843
Qy 389 AspTyrCysAspIleIleLySHisPheMetAspMetSerThrIleLySerLySleuGlu 408
Db 1844 GACTACCATGACATCATTTAAGACCCCATGAGCTCAGCACTCTCAAGCGGAAGATGGAG 1903
Qy 409 AlaArgGluTyrArgAspAlaGlnGluPheGlyAlaAspValAlaArgLeuMetPheSerAsp 428
Db 1904 AACCGTGAATACCGGAGATGACAGAGTTTCTGCTGATATACGGCTTATGTTCTCCAC 1963
Qy 429 CysTyrLySThrAspProAspHisGluValAlaAlaMetAlaArgLySleuGlnAsp 448
Db 1964 TGCTATTAAGTACATCCCCAGATCCAGATGTTGTGGCAATGGCAGAAAGTACAGGAT 2023
Qy 449 ValPheGluMetCysPheAlaLySMetProAspGluProGluGluProValAlaAlaVal 468
Db 2024 GATTGTGAGTTCGCTATGCAAGATGCGAGATGAACCATTAACACAGGCGCTTTACCA 2083
Qy 469 SerSerProAlaValProProPro---ThrLySValAlaValProProSerSerAsp 487
Db 2084 GTCTTACTGCGCATGCGCCCTGCGTGGCCCAATTCGTTCAAGATCTTCACTAGAGAA 2143
Qy 488 SerSerSerAspSerSerSer-----Asp 495
Db 2144 AGTACAGATGAGAGCTCTCTGAGAAAGAGAGAGAGATAGAGAGACGAGAGAA 2203
Qy 496 SerAspSerSerThrAspSerSerGluGluGluArgAlaGlnArgLeuAlaGluLeuGln 515
Db 2204 GAAGAGAGTGAAGCTCAGACTCAGAGAAAGGCTCATCGCTTACAGAACTACAG 2263
Qy 516 GluGlnLeuLySAlaValHisGluGlnLeuAlaLeuSerGlnProGlnGlnAsnLyS 535
Db 2264 GAACAGCTTCGGGCACTACATGAACAACCTGCTGCTGCCAGGGGTCCAAATATCCAG 2323
Qy 536 ProLySlySglMetLySAspLySlySglLySlyS 547
Db 2324 CCCAAGAGCAAAAGAGAAAAA 2359

RESULT 7
LOCUS B0651563 987 bp mRNA linear EST 15-JUL-2002
DEFINITION AGENCOURT 8302486 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6271145
5' mRNA sequence.
ACCESSION B0651563
VERSION B0651563.1 GI:21775735
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 987)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: CGAP (Stanford)

```

```

cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed By: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LNC2447 row: g column: 18
High quality sequence start: 2
High quality sequence stop: 653.
Location/Qualifiers
1..987
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6271145"
/clone_1b="NIH MGC_100"
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="organ: liver; Vector: pOTB7, site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dt priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH-MGC
Library."
BASE COUNT 257 a 315 c 267 g 136 t 12 others
ORIGIN
Alignment Scores:
Pred. No.: 5,6e-78 Length: 987
Score: 1494.00 Matches: 288
Percent Similarity: 96.37% Conservative: 4
Best Local Similarity: 95.05% Mismatches: 11
Query Match: 41.06% Indels: 1
DB: 14 Gaps: 0
US-09-700-590A-22 (1-688) x B0651563 (1-987)
Qy 250 ProProProValProProGlnProGlnProProAlaProAlaProGlnProValGln 269
Db 9 CCGGGGTGGCTGCCCCCGACACACACCCCGCGCTCCAGCTCCAGCCGATACAG 68
Qy 270 SerHisProProIleIleAlaIleThrProGlnProValLySlySlySlyValLyS 289
Db 69 AGCCACCCACCATCATCGCGGCCACCCACAGCTGTGAACAACAAGAGAGATGAAG 128
Qy 290 ArgLySAlaAspThrThrThrProThrThrIleAspProIleHisGluProProSerLeu 309
Db 129 AGGAAAGCAGACACACACACCCCCACCATTCACCCATTCACGAGCCACCTCGCTG 188
Qy 310 ProProGluProLyThrThrLySleuGlyGlnArgArgLySleuSerArgProValLyS 329
Db 189 CCCCCGAGCCCAAGACCAACCAAGCTGGGCGCAGCGGGAAGAGCGCGCTTGAAA 248
Qy 330 ProProLySlyAspValProAspSerGlnGlnHisProAlaProGluLySerSerLyS 349
Db 249 CCTCCAAAGAAAGAGCTGCGGACCTCTACAGCAGACCCAGACCAAGAGAGAGCAAG 308
Qy 350 ValSerGluGlnLeuLySlySlySlySlyYlleuLySglMetPheAlaLySlyHis 369
Db 309 GTCTGGAGCACTCAAGCTGCGAGCGGCATCTCAAGAGATGTTGCCAAGAGCAC 368
Qy 370 AlaAlaTyrAlaTyrProPheTyrLySProValAspValAlaLeuGlyLeuHisAsp 389
Db 369 GCGGCTTACGCTTGCGCTTCTTCAAGCTTGGAAGGAGGACCTGGGCTTACAGAC 428
Qy 390 TyrCysAspIleIleLySHisPheMetAspMetSerThrIleLySerLySleuGluAla 409
Db 429 TACTGTGACATATCAAGACCCCATGACATGACACACATTAATCTTAATGAGGCC 488
Qy 410 ArgGluTyrArgAspAlaGlnGluPheGlyAlaAspValAlaArgLeuMetPheSerAsnLyS 429

```


Db	Accession	Version	Source	Organism	Reference Authors Title Journal	Remark Comment
Db	489	CGTGAAGTACCGTATGCTCAGAGAGTTTGTCGTGCTAGACGTCGACATTTGATGATCTCCACACTGC				
Qy	430	TYRISTYTAENP:ProAspHisGluValAlaMetAlaArgLysLeuGlnAspVal				
Db	549	TATTAAGTACAAACCTCTCGACCATGAGGTGGCCAGCGCCCGCAGCTCCAGAGATGTG				
Qy	450	PheGluMetArgPheAlaIleMetPheProAspGluPProGluGluPProValAlaValSer				
Db	609	TTTCGAAAAGCCCTTTGGCCAGAGATGCGGACGAGCTTGAGAGCCAGCTGTGTGGCGTGTTC				
Qy	470	SerProAlaValProProProThrLysValAlaAlaProProSerSerSerAspSerSer				
Db	669	TCCCGCGCAGTGTCCCTCTCCACCAAGTTGTGGCCCGCCCTCATTCAGGACAGCAGC				
Qy	490	SerAspSerSerSerAspSerSerSerSerSerThrAspAspSerGluGluGlnArgAlaGln				
Db	729	AGCGATGACTCCTCGCAGACAGTACGAGTTGCAGCTGATGACTTNGAGAGACCGAGCCAG				
Qy	510	ArgLeuAlaGluLeuGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu				
Db	789	CGGTGGCTGAGCTCAGAGAGAGCTCAAAAGCGGACGACGACAGCTTGACGCTCTCT				
Qy	530	GlnProGlnGlnAsnLysProLysLysLysGluLysAspLysLysGluLysLysGlu				
Db	848	CAGCCCGCAGACACACACCAAAAAAAAAAAAAAAAAAGGANNNNNNNANANAGAA				
Qy	550	LysHisLys 552				
Db	908	AGAAAAA 916				
RESULT 8						
LOCUS	BC001885					
DEFINITION	BC001885	2440 bp	mRNA	linear	HTC 12-JUL-2001	
VERSION	BC001885					
KEYWORDS	BC001885.1	GI:14706221				
SOURCE	HTC.					
ORGANISM	Homo sapiens.					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.					
TITLE	1 (bases 1 to 2440)					
JOURNAL	Strausberg, R.					
	Direct Submission					
	Submitted (29-JAN-2001) National Institutes of Health, Mammalian					
	Gene Collection (MGC), Cancer Genomics Office, National Cancer					
	Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,					
	USA					
	NIH-MGC Project URL: http://mgc.nci.nih.gov					
	Contact: MGC help desk					
	Email: cgabbs-r@mail.nih.gov					
	Tissue Procurement: Louis M. Staudt, M.D., Ph.D.					
	cDNA Library Preparation: Rubin Laboratory					
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LILNL)					
	DNA Sequencing by: Genome Sequence Centre,					
	BC Cancer Agency, Vancouver, BC, Canada					
	info@bcgsc.bc.ca					
	Steven Jones, Jennifer Asano, Ian Bosdel, Yaron Butterfield,					
	Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,					
	Leticia Hsiao, Martin Krzywnicki, Reta Kutsche, Oliver Lee, Soo					
	Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven					
	Nees, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline					
	Schein, Duane Smalms, Michael Smith, Lorraine Spence, Jeff Stott,					
	Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,					
	George Yang, Scott Zuyderduyn, Marco Marra.					
	Clone distribution: MGC clone distribution information can be found					
	through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov					
	Series: RAW Plate: 4 Row: n Column: 4					
	This clone was selected for full length sequencing because it					
	passed the following selection criteria: matched mRNA gi: 4757809					
	This clone has the following problem: frame shifted.					

FEATURES		Location/Qualifiers
SOURCE	1. 2440	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/clone="IMAGE:3544244"	
	/issue_type="lymph, Burkitt lymphoma"	
	/clone_1ib="NIH MGCC 8"	
	/lab_host="DH10B-R"	
	/note="Vector: pORB7"	
BASE COUNT	619 a 710 c 609 g 502 t	
ORIGIN		
Alignment Scores:		
Pred. No.:	4.06e-77	Length: 2440
Score:	1484.50	Matches: 318
Percent Similarity:	63.028	Conservative: 62
Best Local Similarity:	52.748	Mismatch: 121
Query Match:	40.798	Indels: 103
DB:	11	Gaps: 14
US-03-700-590A-22 (1-688) x BC001895 (1-2440)		
QY	5 serglyProglYThrArgLeuArgAsnLeuProValMetGlyAspGlyLeuGlnThrSer	24
DB	759 GCACGACCCGAGGAAAGAGATTGCAAAACCCCTCTCTGTATGAGAGGGCTTTGAGAGCCCC	818
QY	25 GlnMetSerThrThrGlnAlaGlnAlaGlnProGlnProAlaAsnAlaLaserThrAsn	44
DB	819 ACAATGGCTTGGCGCTGCT--TTGCAACTTACCCCTGCC-----AAC	860
QY	45 ProProProProGlnThrSerAsnProAsnLysProLysArgGlnThrAsnGlnLeuGln	64
DB	861 CCACCAACCCCGGAGGCTGCTCAATCCCAAAAAGCCAGAGCGAGTATCCACACAGCTGCA	920
QY	65 TyrLeuLeuArgValValLeuLysThrLeuTrpLysHisGlnPheAlaTrpProPheGln	84
DB	921 TACCTACCAAGAGTAGATGATGAAGAGCTGTGGAAACATCAGTTGCGATGGCCATTCCGG	980
QY	85 GlnProValAspAlaValLysLeuAsnLeuProAsp-----	96
DB	981 CAGCTGTGATGCTGTCAAACTGGGTCTACCGAT-TCCACCTCGGGTGGAGAGAC	1039
QY	96 -----	96
DB	1040 CACGGTGCCAAAATTTTTCGCTTCTTCCTTCCCTCATGCAAGCCATGATAGCCACC	1099
QY	97 -----TyrTyrLysIleIleLysThrPrometAspMetGlyThrIleLysLysArg	113
DB	1100 CCAAGAGATTATCACAAAATTATTAACACAGCTATGACATGGGTACTTATTAAGAGGAGA	1158
QY	114 LeuGlnLysAsnAsnTyrTyrTPAsnAlaGlnGlnLysIleGlnAspPheAsnThrPhe	133
DB	1160 CTTGAAAACATTTATTTATTTGGCGTCTTCAGATGTATGATCAAGATTATTAATCCATGTC	1218
QY	134 ThrAsnCyTrIleTyrAsnLysProGlyAspAspIleValLeuMetAlaGlnAlaLeu	153
DB	1220 ACCAATCTGTATCATTTTCAACACAGCCACCTGATGATATTTGCTTAATGACAAACGCTG	1279
QY	154 GlyLysLeuPheLeuGlnLysIleAsnGlnLeuProThrGlnGlnThrGlnIleMetIle	173
DB	1280 GAAAGAAATTCCTCAACAGAGGTTGCATCAATGCCAAGAAAGAACAGAGCTGGTAGTG	1338
QY	174 ValGlnAlaLysGly-----ArgGly-----ArgGlyArgLys	184
DB	1340 ACCATCCCTTAAGAACACGCCACAAGAGGGGGCCAAATGGTCAGACGCTCCAGGGCAGTGT	1399
QY	185 GlnThrGlyThrAlaLysProGlyValSerThrValProAsnThrThrGlnAlaSerThr	204
DB	1400 ACCACGTGCCATCAGGTGCTGCGCTCTTCTGTGTATCAACACAGCCCTGATATCTCT	1455
QY	205 ProProGlnThrGlnThrProGlnProAsnProProProValGlnAlaThrProHisPro	224
DB	1460 CCACCTGGAGATA-----CTTACACAGTGTCTCTCAACATTTCCCAACCA	1501

```

QY 225 PheProAlaValThrProAspLeuIleValGlnThrProValMetThrValValProPro 244
DB 1502 -----TCAGTCATTTCTCTCCACTCTCCAGTCCTTGACAC--- 1537
QY 245 GlnProLeuGlnThrProProProValProProGlnProGlnProProAlaProAla 264
DB 1538 -----TCTGTGAGACCCCGCTCTCTGTTACTGACGCTCTCCAGCC--- 1582
QY 265 ProGlnProValGlnSerHisProProIleIleAlaIleThrProGlnProValLysThr 284
DB 1583 -----CAGCCCTTGCCAG 1597
QY 285 LysLysGlyValLysArgLysAlaAspThrThrThrProThrThrIle----- 300
DB 1598 AAAAAAGGCGTTAAAGCGGAAGACAGATCTACACCCCTACACTACAGCAGCATTTGGCT 1657
QY 301 -----AspProIleHisGluProProSerLysLeuProGluProLysThrThrLysLeu 318
DB 1658 CCGGTCTCTCCAGCTAGCCCTCTGGAGTCTT-----GAGCCTAAGCGACAGACGCTT 1711
QY 319 ---GlyLysArgArgLysSerSerArgProValLysProProLysLysAspValProAsp 337
DB 1712 CCCCCTATGCGTGAAGAGAGTGGTGGCCCATCAGCCCCACAGCAAGACTTGCTGAC 1771
QY 338 SerGlnGlnHisProAlaProGluLysSerLysValSerGlnLysGlnLysCysCys 357
DB 1772 TCTCAGACAAACACACAGAGCTCTAAGAAAGAAAGAGCTTACAGAACGTTAAACAATTC 1831
QY 358 SerGlyTlleuLysGlnMetPheAlaLysLysHisAlaAlaTyrAlaTyrProPheTyr 377
DB 1832 AATGGCATTTTGAAGAGATTAATCTCTTAAAGAACATGCTCTCATGCTTGGCTTCTAT 1891
QY 378 LysProValAspValGluAlaLeuGlyLeuHisAspTyrCysAspIleIleLysHisPro 397
DB 1892 AAACCGATGATCTTCTGACCTTGCGCTGACATGACTACATGACATCATTAAGACACCC 1951
QY 398 MetAspMetSerThrTlleLysSerLysLeuGlnAlaArgGlnTyrArgAspAlaGlnGln 417
DB 1952 ATGACACTGACACTGTCAGAGCGGAAGATGAGAACCGTATTAACGGGATGACAGGAG 2011
QY 418 PheGlyAlaAspValArgLeuMetPheSerAsnCysTyrLysTyrAsnProProAspHis 437
DB 2012 TTTCGCTGATGATGAGGCTTATGTTCTCCAACTGCTATTAAGTAACTATCCCAAGATC 2071
QY 438 GlnValValAlaMetAlaArgLysLeuGlnAspValPheGlnMetArgPheAlaLysMet 457
DB 2072 GATGTTGTGGCAATGGACAGAAAGCTACAGATGTAATTGAGTTCCGTTATGCGCAAGAT 2131
QY 458 ProAspGlnProGluGlnProValValAlaValSerSerProAlaValProProPro--- 476
DB 2132 CCGATGAAACCACTAAGAACAGGCGCTTACCACTCTCTACTGCAATGCCCTCGGCTTG 2191
QY 477 ThrLysValValAlaProProSerSerSerSerSerSerSerSerSerSerSerSer--- 494
DB 2192 GCCAAATCGCTTCAAGATCTCCAGTGAAGAAAGTACAGTACAGCTCTCTCGAGAA 2251
QY 495 -----AspSerAspSerSerThrAspAspSerGln 504
DB 2252 GACGAGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2311
QY 505 GlnGluArgAlaGlnArgLeuAlaGlnLeuGlnGlnGlnLeuLysAlaValHisGlnGln 524
DB 2312 GAAGAAAGGCGCTATGCTTACAGCAACTACAGAAACAGCTTGGGCAAGTACATGAACAA 2371
QY 525 LeuAlaAlaLeuSerGlnProGlnGlnAsnLysProLysLysLysGlnLysAspLysLys 544
DB 2372 CTGGCTGCTCTGTCAGAGGTCCAAATATCAAGCCCAAGAGAGAGAGAGAGAGAGAGAG 2431
QY 545 GlnLysLys 547
DB 2432 AAAAAAAA 2440

```

```

RESULT 9
BQ219777 893 bp mRNA linear EST 02-MAY-2002
LOCUS AGENCYCOURT_7644864 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:6020711
DEFINITION 5', mRNA sequence.
ACCESSION BQ219777
VERSION BQ219777.1 GI:20401177
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 893)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/BLNL at:
http://image.llnl.gov
Plate: LLM1325 row: d column: 24
High quality sequence stop: 675.
FEATURES
source location/Qualifiers
1..893
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6020711"
/clone_lib="NIH MGC 70"
/tissue_type="epitheloid carcinoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: pancreas; Vector: pCMV-SPORT6; Site 1: Nct1;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.1 kb. Library constructed by Life
Technologies."
BASE COUNT 259 a 298 c 199 g 137 t
ORIGIN
US-09-700-590A-22 (1-688) x BQ219777 (1-893)
QY 63 LeuGlnTyrLeuLeuArgValValLeuLysThrLeuTyrLysHisGlnPheAlaTyrPro 82
DB 3 CTGCAATACCTGCTGACAGTGGTGCACAGACACTATGAGAAACACAGTTTGATGGCT 62
QY 83 PheGlnGlnProValAspAlaValLysLeuAsnLeuProAspTyrTyrLysIleLys 102
DB 63 TTCACAGAGCTGTGATGATCCGTCAGAGCTGAACCTCCGATTTACTATAGATCATTA 122
QY 103 ThrProMetAspMetGlyThrTlleLysLysArgLeuGlnAsnAsnTyrTyrTrpAsnAla 122
DB 123 AGCCCTATGATATGGGAACAATAAAGAGCGCTTGAGAAACAACATTAATACGGAATGCT 182
QY 123 GlnGluCysIleGlnAspPheAsnThrMetPheThrAsnCysTyrTlleTyrAsnLysPro 142
DB 183 CAGGAATGATCCAGAGACTTCAACTATATGTTTACCAATTTGTTACATTCACAAAGCCT 242
QY 143 GlyAspAspIleValLeuMetAlaGluAlaLeuGlnLysLeuPheLeuGlnLysIleAsn 162
DB 243 GAGATGACATATGCTTAATGGAGAGAGCTCTGAGAAAGCTCTTCTTGCAAAAATTAAT 302
QY 163 GlnLeuProThrGlnGlnThrGlnLysMetIleValGlnAlaLysGlyArgGly 182

```

```

Db      303  GAGCTACCCACAGAAAGAACGAGATCATGATTCAGAGGAAAGAGAGAGCGGG 362
Qy      183  ATGlysgluthrgrlthralalysProglYValserThralProasnthrThGlnAla 202
Db      363  AGGAAAGAAACAGGAGACGAAACCTGGCTTCCACGGACCAACACACACTCAAGCA 422
Qy      203  SerThProProglInthrglnthThProglInProasnthProProglAlaThPro 222
Db      423  TCGACTCTCCCGAGACCCAGACCCCTCAGCCGATCCTCTCTGTCAGAGCGCGCT 482
Qy      223  HisProheProAlaValThProAspleuileValGlnthThProValMetThValVal 242
Db      483  CACCCCTTCCCTGCGTCACCCCGACCTCATGTCACAGACCCCTGTCATGACAGTGTG 542
Qy      243  ProProglInProleuglnthThProProProValProProglInProProProAla 262
Db      543  CTTCCCAAGCCACTGACAGCCCGCCGCACTGCCCCCGACCAACACCCCAACCGCT 602
Qy      263  ProAlaProglInProValGlnSerHisProProleileAlaAlaThProglInProVal 282
Db      603  CCGACTCCCGACCGCGTACAGACCCACCATCATCGCGCCACCCCAAGCTGTG 662
Qy      283  LysThrLysLysGlyValLysArgLysAlaAspThrThrThrProThrThrIleAspPro 302
Db      663  AAGCAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 722
Qy      303  IleHisGlnProProSerLysProProglInProLysThrThrLysLysGly-GlnArgAr 322
Db      723  ATTACAGACCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 782
Qy      322  g-GluserSerArg-ProValLys---ProProLysLysAspValProAspSer---Gln 339
Db      783  GGGAGAGAGAGCGGGCGTGTAAAACTCCCAAGAAAGAGAGAGAGAGAGAGAGAG 842
Qy      340  GlnHisProAlaProGluLysSerSer 348
Db      843  CAGCAGCCAGGACCCAGAGAGAGAGAG 869

RESULT 10
LOCUS   B0214403 933 bp mRNA linear EST 02-MAY-2002
DEFINITION AGENCOURT_7589671 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:6067786
5', mRNA sequence.
ACCESSION B0214403
VERSION   B0214403.1 GI:20395803
KEYWORDS  EST.
SOURCE    human.
ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 933)
           NIH-MGC http://mgi.nci.nih.gov/.
AUTHORS   National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE     Unpublished (1999)
JOURNAL   Contact: Robert Strausberg, Ph.D.
COMMENT   Email: cgabbs-i@mail.nih.gov
           Tissue Procurement: ATCC
           cDNA Library Preparation: Life Technologies, Inc.
           cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
           DNA Sequencing by: Agencourt Bioscience Corporation
           Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LNL at:
           http://image.lnl.gov
           Plate: LHAM3347 row: n column: 11
           High quality sequence stop: 659.
FEATURES
     source
       1..933
       /organism="Homo sapiens"
       /db_xref="taxon:9606"
       /clone="IMAGE:6067786"
       /clone_id="NIH_MGC_92"
       /tissue_type="embryonal carcinoma, cell line"

```

```

BASIC COUNT 288 a 304 c 196 g 145 t
ORIGIN
Alignment Scores:
Pred. No.: 5,46e-73 Length: 933
Score: 1408.00 Matches: 270
Percent Similarity: 90.10% Conservative: 3
Best Local Similarity: 89.11% Mismatches: 21
Query Match: 38,69% Indels: 9
DB: Gaps: 2

US-09-700-590a-22 (1-688) x B0214403 (1-933)
Qy      5  SerGlyProGlyThralArgLeuArgAsnLeuProValMetGlyAspGlyLeuGlnThrSer 24
Db      16  AGCGGCCCTGAGAGAGATGAGAAATCTGCCAGTAATGGGAGATGAGTAAGAACTTCC 75
Qy      25  GlnMetSerThrThrGlnAlaGlnAlaGlnProGlnProAlaAsnAlaIleSerThrAsn 44
Db      76  CAATGTCTACACACAGGCGCCAGCCCAACCCAGCCAGCCCAACGAGCCAGCAGCAAC 135
Qy      45  ProProProProglInthThSerAsnProAsnLysProLysArgGlnthThAsnGlnLeuGln 64
Db      136  CCCCCGCCCCAGAGACTCCCAACCTCAACAGCCCAAGAGGAGAGAGAGAGAGAGAG 195
Qy      65  TyrLeuLeuArgValValLeuLysThrLeuThrLysHisGlnIleAlaThrProPhleGln 84
Db      196  TACCTGCTCAGAGAGGAGGCTCAAGACACTATGGAACACCAAGTTTGATGCGCTTTCAG 255
Qy      85  GlnProValAspAlaValLysLeuAsnLeuProAspTyrTyrLysIleIleLysThrPro 104
Db      256  CAGCCTGTGATGCGCTCAAGAGCTGAGACCTCGATTAATTAATGATCATTTAAAGCGCT 315
Qy      105  MetAspMetGlyThrIleLysLysArgLeuGlnAsnAntyTyrTTPAsnAlaGlnGlu 124
Db      316  ATGATATGGGAACAATAAAGAGCGCTTGAAACACACTATTACTGATGCTCAGAA 375
Qy      125  CysIleGlnAspPheAsnThrMetPheThrAsnGlyTyrIleTyrAsnLysProGlyAsp 144
Db      376  TGTATCCAGAGACTTCAACACTATGTTTACAAATGTTTACATCTTCAACCAAGCCGTGAGAT 435
Qy      145  AspIleValLeuMetAlaGlnAlaLeuGlnLysLeuPheLeuGlnLysIleAsnGlnLeu 164
Db      436  GACATAGCTTAATGGCAGAAAGCTTGGAAGAGCTTCTTTCGAAATAATTAATGAGCTA 495
Qy      165  ProThrGlnGlnThrGlnIleMetIleValGlnAlaLysGlyArgGlyArgLys 184
Db      496  CCCACAGAAAGAAACCGAGATCATGATGTCACAGCAAAAGAGAGAGAGAGAGAGAG 555
Qy      185  GlnThrGlyThralAlysProGlyValLysSerThValProAsnThrThrGlnAlaSerThr 204
Db      556  GAAACAGGAGACAGCAAAACCTGGCGTTTCCACGTTACCAACCACTCAAGCATTCAGCT 615
Qy      205  ProProglInthrglnthThProglInProasnthProProProValGlnAlaThProHisPro 224
Db      616  CTTCCGAGAGACCCAGCCCTCAGCCGAAATCTCTCTGTCAGAGGCGAGCTCACCCC 675
Qy      225  PheProAlaValThProAspleuileValGlnthThProValMetThValValProPro 244
Db      676  TTCCCTGCGGTACACCCCGAGACTCATGTCACAGCCCTGTGACAGAGAGTGTCTCC 735
Qy      245  GlnProleuglnthThProProPro--ValProProglInProglInProProProAlaPro 263
Db      736  CAGCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 795
Qy      264  AlaProglInProValGlnSerHisPro-ProIleIleAlaAlaThProglInProVal-- 282

```

DB 796 GCTCCCAAGCCCAAGAGACCCCAACCATCATTCGCGGCGCACCCCAAGGCGCTG 855

QY 283 -----LysThrLysGlyValLysArgLysAlaSerThrThr 296

DB 856 GGGAGAGACCAAAAAAGGAGAGTGGAAAAAGGAAAAAGCCCAAAATCCAAACCAATCC 915

QY 296 rProthr 298

DB 916 CCCACC 922

RESULT 11

LOCUS B0646784 951 bp mRNA linear EST 15-JUL-2002

DEFINITION AGENCOURT_8298482 NIH_MGC_100 Homo sapiens CDNA clone IMAGE:6270294

ACCESSION B0646784

VERSION B0646784.1 GI:21770956

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 951)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: CGAP (Stanford)
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>
Plate: LNCM2445 row: d column: 07
High quality sequence stop: 636.
Location/Qualifiers

FEATURES

source 1..951

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:6270294"

/clone_lib="NIH_MGC_100"

/issue_type="hepatocellular carcinoma, cell line"

/lab_host="DH10B (phage-resistant)"

/note="Organ: liver; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC library."

BASE COUNT 271 a 305 c 226 g 148 t 1 others

ORIGIN

Alignment Scores:

Pred. No.: 5.54e-73 Length: 951

Score: 1408.00 Matches: 271

Percent Similarity: 94.83% Conservative: 4

Best Local Similarity: 93.45% Mismatches: 11

Query Match: 38.69% Indels: 4

DB: 14 Gaps: 0

US-09-700-590A-22 (1-688) x B0646784 (1-951)

QY 1 MetSerAlaGluSerGlyProGlyThrArgLeuArgAsnLeuProValMetGlyAspGly 20

DB 89 ATGTCTCGGAGAGCGCCCTGGGACAGATTGGAATCTGCCAGTAATGGGGGATGGA 148

QY 21 LeuGluThrSerGlnMetSerThrThrGlnAlaGlnAlaGlnProGlnProAlaAsnAla 40

DB 149 CTAGAAACTTCCCAATGTCTACACACAGGCCCAAGCCCAACCCCAAGCCAGCAACGCA 208

QY 41 AlaSerThrAsnProProProGluThrThrSerAsnProAsnLysProLysArgGlnThr 60

DB 209 GCCAGACCAACCAACCCCGCCGCCAGAGACCTCCAACTTAACAGCCCAAGAGGAGAGCC 268

QY 61 AsnGlnLeuGlnThrLeuLeuArgValValLeuLysThrLeuThrLysHisGlnPheAla 80

DB 269 AACCACTGCATACCTGCTCAGAGTGTCTCAGACACTATGGAACCAACGATTTGCA 328

QY 81 TrpProPheGlnGlnProValAlaValLysLeuAsnLeuProAspTyrTyrLysIle 100

DB 329 TGGCTTTCCAGCAGCGCTGTGATGCCGTCAAGCTCAACCTCCCTATTACTTAAGATC 388

QY 101 IleLysThrProMetAspMetGlyThrIleLysLysArgLeuGlnAsnAsnTyrTyrTrp 120

DB 389 ATTAACACCGCTATGATATGGGAACAATAAGAACCGCTGGAAAAACAATTAATCTG 448

QY 121 AsnAlaGlnGluCysIleGlnAspPheAsnThrMetPheThrAsnCystrIleTyrAsn 140

DB 449 AATGCTCAGGAATGTATCCAGGACCTTCAACACTATGTTTAAATTTGTTACATCTACAAC 508

QY 141 LysProGlyAspAspIleValLeuMetAlaGluAlaLeuGlnLysLeuPheLeuGlnLys 160

DB 509 AAGCTCGAGATGACATGACTTAAATGGCAGAAAGCTGTGAAAAAGCTCTTTCGAAAAA 568

QY 161 IleAsnGlnLeuProThrGluGluThrGluIleMetIleValGlnAlaLysGlyArgGly 180

DB 569 ATTAATGAGTACCCACAGAAAGAACCGAGATCATATATGTCAGGCAAAAGAAAGGA 628

QY 181 ArgGlyArgLysGluThrGlyThrAlaLysProGlyValSerThrValProAsnThrThr 200

DB 629 CTGGGAGAGAAAGAAACAGAGACAGCAAAACCTGGCGTTTCCACGGTACCAACACAACT 688

QY 201 GlnAlaSerThrProProGlnThrGlnThrProGlnProAsnPro-ProProValGlnAl 220

DB 689 CAAGCATGCACTCTCCGAGACCCGAGACCCCTCAGCCCAATCTCTCTGTGAGAGC 748

QY 220 aThrProHisProPheProAlaValAlaThrProAspLeuIleValGlnThrProValMet 240

DB 749 CAGGCTCACCCCTTCTCCGTGGGTACCCCGAGACCTCATGTCACACCCCTGTCTATGA 808

QY 240 rValValProProGlnProLeuGln-ThrProProProVal-ProProGlnProGlnPro 259

DB 809 CGTGGTCCCTCCCAACCCACTGAGAAAGCCCGCCAGTGGGCCCCGACAGCAACACC 868

QY 260 ProProAlaProAlaProGlnProValGlnSerHisProProIleIleAla-AlaThrTr 279

DB 869 CCACCGGCTTCAGGCTCCCAAGCCCGGAAAGAGCCACCCCAATTTTCGCGGGCCACCCC 928

QY 279 oglnProValLysThrLysLys 286

DB 929 ACAGCCCGGAGAAACCAAGAA 950

RESULT 12

LOCUS B0645818 925 bp mRNA linear EST 15-JUL-2002

DEFINITION AGENCOURT_8350724 NIH_MGC_100 Homo sapiens CDNA clone IMAGE:6287189

ACCESSION B0645818

VERSION B0645818.1 GI:21769990

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 925)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: CGAP (Stanford)

CDNA Library Preparation: Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
 Plate: LUCM2489 row: d column: 06
 High quality sequence stop: 684.

FEATURES

source
 1. .925
 Location/Qualifiers

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:6287189"
 /clone_id="NIH_MGC_100"
 /rname="hepatocellular carcinoma, cell line"
 /note="Organ: liver; Vector: pOTB1; Site_1: XhoI; Site_2:
 EcoRI; CDNA made by oligo-dT priming. Directionally cloned
 into EcoRI/XhoI sites using the following 5' adaptor:
 GGCAAGAG(G). Size-selected >500bp for average insert size
 1.8kb. Library constructed by Ling Hong in the laboratory
 of Gerald M. Rubin (University of California, Berkeley)
 using ZAP-CDNA synthesis kit (Stratagene) and Superscript
 II RT (Life Technologies). Note: this is a NIH_MGC
 library."

BASE COUNT 261 a 295 c 220 g 147 t 2 others

ORIGIN

Alignment Scores:

Pred. No.: 4.73e-70 Length: 925
 Score: 1357.50 Matches: 264
 Percent Similarity: 93.33% Conservative: 2
 Best Local Similarity: 92.63% Mismatches: 11
 Query Match: 37.30% Indels: 9
 DB: 14 Gaps: 3

US-09-700-590a-22 (1-688) x B0645818 (1-925)

Qy 1 MetSerAlaGluSerGlyProGlyThrArgLeuArgAnLeuProValMetGlyAspGly 20
 Db 89 ATGCTCGGAGAGAGGGCCCTGGAGCGAGATTGAGAAATCTCCAGTAAATGGGGATGA 148
 Qy 21 LeuGluThrSerGlnMetSerThrThrGlnAlaGlnAlaGlnProGlnProAlaAsnAla 40
 Db 149 CTAGAACTTCCCAATGATCTCAACACAGGGCCAGCCCAACCCAGCCAGCCCAACCA 208
 Qy 41 AlaSerThrAnProProProProGluThrSerAnProAnLysProLysArgGlnThr 60
 Db 209 GCCAGCACCAACCCCGCCGCCAGAGACCTTCAACACCCCAAGGCGAGACC 268
 Qy 61 AsnGlnLeuGlnThrLeuLeuArgValValLeuLysThrLeuTrpLysHisGlnPheAla 80
 Db 269 AACCACTGCAATACCTCTGTCAGAGTGTGCTCAAGACCTATGAGAAACACAGTTTGA 328
 Qy 81 TrpProPheGlnGlnProValAlaValLysLeuAnLeuProAspTyrTyrLysIle 100
 Db 329 TGGCTTTCCAGCAGACCTGTGATGCCGTCAAGCTGAACTCCCTGATTAATAGATC 388
 Qy 101 IleLysThrProMetAspMetGlyThrIleLysLysArgLeuGlnAsnAsnTyrTyrTrp 120
 Db 389 ATTAAAGCGCTATGATATGGAGCAATTAAGAAAGCGCTTGAGAAAAACAATATTACTCG 448
 Qy 121 AsnAlaGlnGlnCysIleGlnAspPheAnThrMetPheThrAsnCysTyrIleTyrAn 140
 Db 449 AATGCTCGAATGATATCGAGACTTCAACACTATGTTTACAAATTTGTTACATCAACAC 508
 Qy 141 LysProGlyAspAspIleValIleuMetAlaGlnAlaLeuGlnLysLeuPheLeuGlnLys 160
 Db 509 AAGCTGAGATGATCATGTCTTATGGAGAGAGCTCTGGAAAAAAGCTCTTCTTGAAAAA 568
 Qy 161 IleAnGlnLeuProThrGlnGluThrGluIleMetIleValGlnAlaLysGlyArgGly 180
 Db 569 ATAAATGAGTACCAAGAGAAAGCAAGATCATGATATGTCAGGCAAAAGAGAGAGA 628

Qy 181 ArgGlyArgLysGlnThrGlyThrAlaLysProGlyValSerThrValProAnThrThr 200
 Db 629 CGTGGAGGAAAGAAACAGAGGACGAGAAACCTGGCTTTCCAGTACCAACACAACT 688
 Qy 201 GlnAlaSerThrProProGlnThrGlnThrProGlnProAnProProProValGlnAla 220
 Db 689 CAACATTCGATCTCTCCGAGAGACCAACCTTCAAGCGAATCTCTCTGTCAGAGCC 748
 Qy 221 ThrProHisProPheProAlaValThrProAspLeuIleValGlnThrProValMetThr 240
 Db 749 ACGCTCAACCTTCTCTGCGGTACCCCGGACTTATCTTCAGACCCCTGTAGAGACA 808
 Qy 241 -ValValProProGlnProLeuGlnThrProPro--ProValProProGlnProGlnPr 259
 Db 809 AGTGCTGCTCCANNCCCTGAGAGCGCCCGGAGGGGCCCCCAACACCAACC 868
 Qy 259 oProProAlaProAlaProGlnProValGlnSerHisPro--ProIleIleAlaIaTh 278
 Db 869 C-----CCAAC-CCGTTTCAGGGTTCCTCCAGGCGGGATTAAGAAAGCCCC 912
 Qy 278 rProGlnProVal 282
 Db 913 CCCACCCCTTTA 925

RESULT 13

B0653187

LOCUS

DEFINITION

AGENCOURT_8341467 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6267841

5', mRNA sequence.

ACCESSION

B0653187

VERSION

B0653187.1 GI:21777359

KEYWORDS

EST.

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 975)

AUTHORS

NIH-MGC <http://mgs.nci.nih.gov/>.

TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgabs-r@mail.nih.gov

Tissue Procurement: CGAP (Stanford)

CDNA Library Preparation: Rubin Laboratory

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:

<http://image.llnl.gov>

Plate: LUCM2438 row: n column: 02

High quality sequence stop: 656.

Location/Qualifiers

1. .975

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:6267841"

/clone_id="NIH_MGC_100"

/rname="hepatocellular carcinoma, cell line"

/lab_host="DH10B (phage-resistant)"

/note="Organ: liver; Vector: pOTB1; Site_1: XhoI; Site_2:
 EcoRI; CDNA made by oligo-dT priming. Directionally cloned
 into EcoRI/XhoI sites using the following 5' adaptor:
 GGCAAGAG(G). Size-selected >500bp for average insert size
 1.8kb. Library constructed by Ling Hong in the laboratory
 of Gerald M. Rubin (University of California, Berkeley)
 using ZAP-CDNA synthesis kit (Stratagene) and Superscript
 II RT (Life Technologies). Note: this is a NIH_MGC
 library."

BASE COUNT 275 a 304 c 246 g 145 t 5 others

ORIGIN

Alignment Scores:

275 a 304 c 246 g 145 t 5 others

Pred. No.: 1.03e-69 Length: 975
 Score: 1352.00 Matches: 263
 Percent Similarity: 90.91% Conservative: 7
 Best Local Similarity: 88.55% Mismatches: 23
 Query Match: 37.15% Indels: 4
 DB: 14 Gaps: 2

US-09-700-590a-22 (1-688) x BQ653187 (1-975)

QY 1 MetSerAlaGluSerGlyProGlyThrArgLeuArgLeuProValMetGlyAspGly 20
 DB 89 ATGCTCGGAGAGAGCGGCGCTGGGACGAGATTGGAATCTGCCAGTAATGGGAGTGA 148
 QY 21 LeuGluThrSerGlnMetSerThrThrGlnAlaGlnAlaGlnProGlnProAlaAsnAla 40
 DB 149 CTGAAACCTTCCCAATATGTCTACACAGGCGCCAGGCCCAACCCAGCCAGCCAGCA 208
 QY 41 AlaSerThrAsnProProProGlnThrSerAsnProAsnProGlySarGlnThr 60
 DB 209 GCCAGCACCAACCCCGCCCGGAGAGACCTCAACCTTAACAGCCCAAGAGGCAACC 268
 QY 61 AsnGlnLeuGlnThrLeuLeuArgValValLeuThrLeuThrPlySH1SGlnPheAla 80
 DB 269 AACCACTGCAATTCCTGCTCAGAGTGCTCTCAAGACATATGAAACACAGTTTGA 328
 QY 81 TrpProheGlnGlnProValAspAlaValLysLeuAsnLeuProAspTyrTyrLys1le 100
 DB 329 TGGCTTTCCAGAGCGCTGTGATGCGCTCAAGCTGAACCTCCTGATTAATTAAGATC 388
 QY 101 IleLysThrProMetLaspMetGlyThrIleLysLysArgLeuGlnAsnGlyTyrTyr 120
 DB 389 ATTAAACCGCTTGTGATATGGAAACATTAAGAGGCTTGGAAAACACTATTACTGG 448
 QY 121 AsnAlaGlnGluCysIleGlnAspPheAsnThrMetPheThrAsnCysTyrIleTyrAsn 140
 DB 449 AATGCTAGAGATATATCCAGAGCTTCAACATATGTTTACAAATTTGTTACATTTAACA 508
 QY 141 LysProGlyAspAspIleValLeuMetAlaGlnAlaLeuGlnLysLeuPheLeuGlnLys 160
 DB 509 AAGCTCGAGATGACATATGCTTAATGCGAGAACCTTGGAAAAAGCTCTTTCGAAAAA 568
 QY 161 IleAsnGluLeuProThrThrGluThrGluIleMetIleValGlnAlaLysGlyArgGly 180
 DB 569 ATAAATAGCTACCCACAGAAAGAAACGAGATCATGATAGCCAGGCAAAAGAAAGAGA 628
 QY 181 ArgGlyArgLysGluThrGlyThrAlaLysProGlyValSerThrValProAsnThrThr 200
 DB 629 CGTGGAGAAAGAAAGAGAGAGACGACAAACCTGGCGTTTCACGGTACCAAAACACA 688
 QY 201 GlnAlaSerThrProProGlnThrGlnThrProGlnProAsnProProProValGlnAla 220
 DB 689 CAAAGCATGACCTCTCCGAGAACCCAGACCCCTCAGCGAATCTCTCTGTCAGAGCC 748
 QY 221 ThrProHisProPheProAlaValThrProAspLeuIleValGlnThrProValMet-Th 240
 DB 749 ACCGCTACACCTTCTCTGCGTACCCCGGACCTCATCTGTCAGACCCCTGGCATGAC 808
 QY 240 rValValProGlnProLeu---GlnThrProProProValProProGlnProGlnPr 259
 DB 809 AGTGTGTCTCCGNNNNCACTTGGCAAGGCCCGCCGCGCAGGGGGCCCCCAAGCACAC 868
 QY 259 oProPro---AlaProAlaProGlnProValGlnSerHisPro-ProIleIleAlaIat 278
 DB 869 CCCACCCCGGTCCAGGTCGCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 928
 QY 278 hrProGlnProValLysThrLysLysGlyValLysArgLysAlaAsp 293
 DB 929 CCCCCCAGGCGCTGGGAGAGAAAAAAGGGAGCGGAAAGGAA 975

RESULT 14 BQ749323 815 bp mRNA linear EST 17-JUL-2002
 LOCUS BQ749323
 DEFINITION UI-M-FCO-byb-e-03-0-UI.r1 NIH_BMAP_FCO Mus musculus cDNA clone

IMAGE:5715818 5', mRNA sequence.
 BQ749323
 BQ749323.1 GI:21896110
 EST.
 KEYWORDS house mouse.
 SOURCE Mus musculus
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 815)
 NIH-MGC <http://img.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: Dr. James Lin, University of Iowa
 cDNA library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/BLN at:
<http://image.llnl.gov>
 This clone was contributed by the Brain Molecular Anatomy Project (BMAP)
 Seq primer: pyx-5.
 Location/Qualifiers
 1. 815
 /organism="Mus musculus"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAG:5715818"
 /clone_1ib="NIH_BMAP_FCO"
 /tissue_type="Whole Brain"
 /dev_stage="embryo 12.5 dpc"
 /lab_host="DH10B (71 phage resistant)"
 /note="Organ: brain; Vector: pyx-acc; Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaldi, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured mRNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with Not I, and then cloned directionally into pyx-acc vector. The library tag sequence located between the Not I site and the polyA tail is TGAAGAGGCC. This library was created for the University of Iowa Mouse Brain Molecular Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institutes of Mental Health (NIMH), Hemlin Chin, Ph.D., program coordinator."

BASE COUNT 221 a 269 c 189 g 136 t
 ORIGIN

Alignment Scores:
 Pred. No.: 7.16e-69 Length: 815
 Score: 1336.50 Matches: 252
 Percent Similarity: 94.10% Conservative: 3
 Best Local Similarity: 92.99% Mismatches: 15
 Query Match: 36.73% Indels: 2
 DB: 14 Gaps: 1

US-09-700-590a-22 (1-688) x BQ749323 (1-815)

QY 170 GluIleMetIleValGlnAlaLysGlyArgGlyArgLysGluThrAla 189
 DB 2 GAGATCATGATATCTCCAGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 61
 QY 190 LysProGlyValSerThrValProAsnThrThrGlnAlaSerThrProProGlnThrGln 209
 DB 62 AAGCTGTGTATTCACGAGAACACACACATCAAGATCACTCTCCGAGACCCAG 121
 QY 210 ThrProGlnProAsn---ProProProValGlnAlaThrProHisProPheProAlaVal 228
 DB 122 ACCGCTCAGCAACCTCTCTCAGCTGTCAGGCGCACAATCACCCTTCTGCTGTC 181

QY 229 ThrProAspleuileValGlnThrProValMetThrValProProGlnProLeuGln 248
 DB 182 ACCCGAGACCTCATGGCCAGACCTCTGTCAAGCAATGGAGGCCCTCCAGCACTTCAG 241
 QY 249 ThrProProProValProProGlnProGlnProProProValProValProVal 268
 DB 242 ACTCTTCACCGGTACCCCTCCAGCACCCCTCCAGCTGTCTCCAGCTGTGTG 301
 QY 269 GlnSerHisProProPolelleAlaAlaThrProGlnProValLysThrLysGlyVal 288
 DB 302 CAGAGTCACCCCGCATATGTGACACCCCGCTGTGAGACAAAGAGGGGTG 361
 QY 289 LysArgLysAlaAspThrThrThrProThrThrLysAspProIleLysGluProProSer 308
 DB 362 AAGAGGAAGAGATACCAACCCCTTACCAACATCGACCCCATTCATGAGCCACCTCA 421
 QY 309 LeuProProGluProLysThrThrLysLeuGlyGlnArgGluSerSerThrProVal 328
 DB 422 CTGGCCCGAGACCCCAAGACCGCAAGCTGGTCTCTCGCGGAGAGAGAGACCTGTG 481
 QY 329 LysProProLysLysAspValProAspSerGlnGlnHisProAlaProGluLysSerSer 348
 DB 482 AAGCTCCAAAGAGATGTACCGACTCAAGACGACCCAGGGCCAGAGAGACAC 541
 QY 349 LysValSerGluGlnLeuLysCysCysSerGlyLysLeuLysGluMetPheAlaLysLys 368
 DB 542 AAGATCTCTGACGACCTAAAGTGTGAGCTGAGCTCTCAAGAGAGATTTTCCAGAAA 601
 QY 369 HisAlaAlaLysAlaLysProPheThrLysProValLysValLysGlyLysHis 388
 DB 602 CATGCTGCTATGCTGCTGCTCTTCTACCAAGCTGTGATGTGAGGCACTGGTCTGAC 661
 QY 389 AspThrCysAspIlelleLysHisProMetAspMetSerThrLysSerLysLeuGlu 408
 DB 662 GACTACGTGACATCATCAATCAATCCATGACATGACATGACATCAATCAATCAATGAG 721
 QY 409 AlaArgGluThrArgAspAlaGlnGluPheGlyAlaAspValArgLeuMetPheSerAsn 428
 DB 722 TCCGAGAGTACAGAGATGCCAGGAATTTGGTGTGATGTCCATGTATGTCTCAAC 781
 QY 429 CysThrLysThrAsnProProAspHisGluVal 439
 DB 782 TGCTACAGATACAC-CCCCCTGACCATGAGTG 813
 RESULT 15
 LOCUS B0650970 926 bp mRNA linear EST 15-JUL-2002
 DEFINITION AGENCOURT 8488559 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6296078
 ACCESSION B0650970
 VERSION B0650970.1 GI:21775142
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 926)
 AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: CGAP (Stanford)
 CDNA Library Preparation: Rubin Laboratory
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
 Plate: LUCM2503 row: f column: 15
 High quality sequence stop: 595.
 Location/Qualifiers

source 1..926
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:6296078"
 /clone_lib="NIH_MGC_100"
 /issue_type="hepatocellular carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: liver; Vector: pOTB7, Site_1: XhoI, Site_2:
 EcoRI; cDNA made by oligo-dT priming. Directionally cloned
 into EcoRI/XhoI sites using the following 5' adaptor:
 GGCACGAG(G). Size-selected >500bp for average insert size
 1.8kb. Library constructed by Ling Hong in the laboratory
 of Gerald M. Rubin (University of California, Berkeley)
 using ZAP-cDNA synthesis kit (Stratagene) and Superscript
 II RT (Life Technologies). Note: this is a NIH_MGC
 Library."
 BASE COUNT 267 a 290 c 218 g 151 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 7..73e-68 Length: 926
 Score: 1319.50 Matches: 256
 Percent Similarity: 92.45% Conservative: 1
 Best Local Similarity: 92.09% Mismatches: 16
 Query Match: 36.26% Indels: 6
 DB: 14 Gaps: 2
 US-09-700-590a-22 (1-688) x B0650970 (1-926)
 QY 1 MetSerLysGluSerGlyProGlyThrArgLeuArgAsnLeuProValMetGlyAspGly 20
 DB 89 ATGTCTCGGAGACCGGCTGGAGACAGATGTAAGATCTGCCAGTAAGGGGAGTGA 148
 QY 21 LeuGluThrSerGlnMetSerThrThrGlnAlaGlnAlaGlnProGlnProAlaAsnAla 40
 DB 149 CTAGAACTTCCAAATGTCTTCAACACAGGCCCAAGCCCAAGCCAGCAAGCA 208
 QY 41 AlaSerThrAsnProProProGluThrSerAsnProAsnLysProLysArgGlnThr 60
 DB 209 GCCAGACCAACCCCGCCGCCAGAGACTCCAACTCAACAGCCCAAGAGGCAAGCC 268
 QY 61 AsnGlnLeuGlnThrLeuLeuArgValLeuLysThrLeuThrPheLysGlnPheAla 80
 DB 269 AACCACTGCAATACCTGCTCAGAGTGCTCTCAAGCACTATGAGAAACCAAGTTTGA 328
 QY 81 ThrProPheGlnGlnProValAspAlaValLysLeuAsnLeuProAspThrLysLysIle 100
 DB 329 TGCCCTTCCAGCAGCTGTGTGATGCCGTCAAGCTGAACCTCCTGATTAATAAGATC 388
 QY 101 IleLysThrProMetAspMetGlyThrIleLysLysArgLeuGluAsnAsnThrLysTrp 120
 DB 389 ATTAAGAGCCTATGATATGAGAAACAATAAGAAAGCGCTTGGAACAACTATTACTGG 448
 QY 121 AsnAlaGlnGluCysIleGlnAspPheAsnThrMetPheThrAsnCysThrIleThrAsn 140
 DB 449 AATCTCAGGAATATATCCAGGACTTCACACTATGTTTCAAAATTTTATCATCTACAC 508
 QY 141 LysProGlyAspAspIleValLeuMetArgLysLeuGluLysLeuPheLeuGlnLys 160
 DB 509 AAGCTGAGATGACATAGTCTTAATGACAGACCTGTGAAAAGCTCTTCTTCAAAAA 568
 QY 161 IleAsnGluLeuProThrArgLysGluThrGlnIleMetIleValGlnAlaLysGlyArgGly 180
 DB 569 ATAATAGAGTACCAACAGAGAAACGAGATCATGATGTCAGGCAAAAGAGAGGA 628
 QY 181 ArgGlyArgLysGluThrArgLysThrAlaLysProGlyValSerThrValProAsnThrThr 200
 DB 629 CGTGGAGAGAAAGAGAGAGAGCAAACTGGCTTTTACAGGTAACCAAAACAACACT 688
 QY 201 GlnLysSerThrProProGlnThrGlnThrProGlnProAsnProProProValGlnAla 220
 DB 689 CAAGCATGACTCCCTCGACAGACCAAGCCCTCAGCGAATCTCTCTGTGACAGGC 748


```
Oy 221 ThrProHisProPheProAlaValThrProAspLeuIleValGlnThrProValMet-Th 240
Db 749 ACSCCTCACCCCTTCTCTGCGGTACCCCGACCTCATGTCAGAACCCCTGCATGAC 808
Oy 240 rValValProProGlnProLeuGlnThr---ProProProValProProGlnProGlnPr 259
Db 809 AGTGTGCTCTCCACACCCCATGAAAAAGGCCCCCGCCAGAGCCCCCAGGTCAAAA 868
Oy 259 oProProAlaPro-----AlaProGlnProValGlnSerHisProPro 273
Db 869 ACCCCCAAC-CCGTTCAAGTTCCCCAGGCCCGGAACAGACCAACCCCCC 919
```

Search completed: March 14, 2003, 22:42:08
Job time : 2141 secs

THIS PAGE BLANK (USPTO)